

Uniwersytet im. Adama Mickiewicza
Wydział Matematyki i Informatyki

Wykłady z Rachunku Prawdopodobieństwa 2

Poznań, 2008/2009

Spis treści

Wstęp	3
1. Wykład 1: Prosty spacer losowy z barierami	4
Wstęp	4
1.1. Prosty spacer losowy	4
1.2. Bariery pochłaniające	6
1.3. Bariery odpychające	8
2. Wykład 2: PSL – metoda zliczania ścieżek	9
Wstęp	9
2.1. Zasada odbicia	9
2.2. Powroty do osi	11
2.3. Zagadnienie pierwszych odwiedzin	13
2.3.1. Technika odwracania	14
3. Wykład 3: Funkcje tworzące	16
Dokończenie poprzedniego wykładu	16
3.1. Funkcje tworzące	17
3.1.1. Liczenie momentów	18
3.1.2. Sumy zmiennych niezależnych	19
3.1.3. Zastosowanie dla spacerów losowych	20
4. Wykład 4: Funkcje tworzące. Procesy gałązkowe	22
Wprowadzenie	22
4.1. Funkcje tworzące (dokończenie)	22
4.1.1. Moment pierwszych odwiedzin	22
4.1.2. Czas przebywania na dodatniej półosi	25
4.2. Procesy gałązkowe	27
5. Wykład 5: Procesy gałązkowe. Łańcuchy Markowa.	29
Wprowadzenie	29
5.1. Procesy gałązkowe (dokończenie)	29
5.1.1. Geometryczny proces gałązkowy	29
5.2. Łańcuchy Markowa	33
6. Wykład 6: Klasyfikacja łańcuchów Markowa. Rozkłady stacjonarne.	36
Wstęp	36
6.1. Rozkład początkowy	36
6.2. Przykłady	38
6.3. Klasyfikacja łańcuchów Markowa	39
6.4. Rozkład stacjonarny	43
7. Wykład 7: Rozkłady stacjonarne. Coupling.	45
Wstęp	45
7.1. Postać rozkładu stacjonarnego	48
7.2. Coupling	48

8. Wykład 8: Zbieżność poissonowska. Twierdzenie ergodyczne. Łańcuchy odwracalne.	50
Wstęp	50
8.1. Dokończenie poprzedniego wykładu	50
8.2. Odwracalne Łańcuchy Markowa	52
9. Wykład 9: Markov Chain Monte Carlo	56
Wstęp	56
9.1. Idea MCMC	56
9.2. Przykłady	56
9.3. Przybliżone przeliczanie	59
10. Wykład 10 i 11: Aproksymacja liczby kolorowań grafu	61
Wstęp	61
10.1. Dowód istnienia algorytmu RPTAS dla obliczania liczby q -kolorowań grafu	61
10.2. Poprawienie oszacowania z Twierdzenia 10.6	67
11. Wykład 12: Martyngały: definicja, twierdzenia o zbieżności	69
Wstęp	69
11.1. Warunkowa wartość oczekiwana	69
11.1.1. Przypadek jednowymiarowy	69
11.1.2. Przypadek wielowymiarowy	70
11.2. Martyngały, definicja i przykłady	71
11.3. Twierdzenia o zbieżności martyngałów	73
12. Wykład 12: Martyngały: nierówność Hoeffdinga	75
Wstęp	75
12.1. Dowody twierdzeń z poprzedniego wykładu	75
12.2. Nierówność Hoeffdinga	77
13. Appendix	80
A - Zadania domowe	80

Wstęp

Niniejszy dokument jest zebraniem wykładów z Rachunku Prawdopodobieństwa 2 odbywających się od października 2008 do stycznia 2009 na Wydziale Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu.

Wykład poruszał zagadnienia procesów losowych oraz ich zastosowań. W trakcie wykładu przedstawione zostały zagadnienia spacerów losowych, wykorzystania funkcji tworzących do procesów losowych, procesów gałęzkowych, łańcuchów Markowa oraz ich wykorzystaniu w algorytmach typu Monte Carlo, a także martyngałów.

Wykład prowadzony był przez prof. Andrzeja Rucińskiego. Treść została spisana przez studentów i doktorantów – uczestników wykładów. Osobami zaangażowanymi w powstanie tego opracowania były (w kolejności spisywania wykładów):

- Łukasz Michniewicz, Magdalena Musioł
- Marek Kaluba, Bartosz Naskręcki
- Krzysztof Krzywdziński, Paweł Wawrzyniak
- Tomasz Ciaś, Paweł Skórzewski
- Łukasz Witkowski, Marcin Witkowski
- Katarzyna Mieczkowska, Bartosz Zaleski
- Arkadiusz Buchelt, Przemysław Sokołowski
- Przemysław Rogowski, Tomasz Rzędowski
- Ewelina Rychlińska, Wojciech Wawrzyniak
- Adam Przestacki, Patryk Szczęśniewski
- Mirosława Jańczak
- Łukasz Waszak
- złożenie i korekta – Rafał Witkowski
- pomoc przy korekcie – Bartosz Naskręcki, Ewelina Rychlińska
- opieka merytoryczna – Andrzej Ruciński

1. Wykład 1: Prosty spacer losowy z barierami

Wstęp

Wyobraźmy sobie osobnika namiętnie oddającego się grze w rzut monetą. Za każdego wyrzuconego orła kasyno wypłaca mu \$1, natomiast za każdą wyrzuconą reszkę musi oddać \$1. Osobnik ten jest typem spod wyjątkowo ciemnej gwiazdy, dlatego też hazard nie jest jego jedynym nałogiem. Rzucając po raz kolejny monetą lubi topić swoją frustrację i nerwy w alkoholu, dlatego rzadko kiedy wraca do domu trzeźwy. Ze względu na swój stan wracając wykonuje losowo krok do przodu lub do tyłu z pewnymi prawdopodobieństwami (zakładamy, że możliwy jest powrót do domu po linii prostej – chodzenie w stanie upojenia alkoholowego jest i tak wystarczająco trudne). Będziemy zastanawiać się między innymi nad tym jak prawdopodobne jest, że przegra on wszystkie przyniesione do kasyna pieniądze. Ponadto przekonamy się, że w pewnych okolicznościach, opuściwszy kasyno w stanie bachicznego natchnienia, nasz bohater najprawdopodobniej do niego powróci.

1.1. Prosty spacer losowy

Podczas tego wykładu będziemy zakładać, że X_1, X_2, \dots jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie prawdopodobieństwa (zmiennie takie oznaczamy w skrócie *iid*) takim, że dla każdego $i \in \mathbb{N}$

$$X_i = \begin{cases} 1, & \text{z prawdopodobieństwem } p \\ -1, & \text{z prawdopodobieństwem } q = 1 - p. \end{cases}$$

Definicja 1.1. *Prostym spacerem losowym* (oznaczanym *PSL*) nazywamy ciąg zmiennych losowych

$$S_n = S_0 + \sum_{i=1}^n X_i,$$

gdzie S_0 jest pewną stałą.

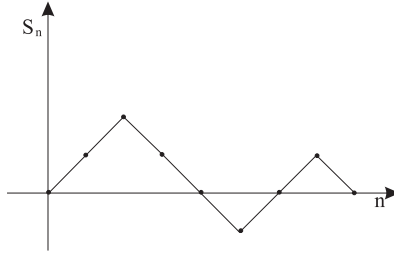
Szczególny przypadek *PSL* dla $p = q = \frac{1}{2}$ nazywamy *symetrycznym PSL* (Pearson).

Wniosek 1.2. Jako oczywisty wniosek z definicji otrzymujemy następujący wzór rekurencyjny:

$$S_{n+1} = S_n + X_{n+1}.$$

Spacer losowy wygodnie jest reprezentować w postaci wykresu w dwuwymiarowym układzie współrzędnych. Oś odciętych oznacza wówczas „czas”, natomiast oś rzędnych „pozycję”

w spacerze losowym (Rysunek 1.1). Wykres taki będziemy nazywać *trajektorią spaceru losowego*.



Rysunek 1.1. Trajektorja spaceru losowego.

Twierdzenie 1.3 (Własności *PSL*).

Niech S_n będzie *PSL*. Zachodzą wówczas następujące własności:

1. *Jednorodność przestrzenna.*

$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}(S_n = j + b | S_0 = a + b)$$

2. *Jednorodność czasowa.*

$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}(S_{m+n} = j | S_m = a)$$

3. *Własność Markowa.*

$$\mathbb{P}(S_{m+n} = j | S_0, S_1, \dots, S_m) = \mathbb{P}(S_{m+n} = j | S_m)$$

Dowód.

1.
$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}\left(\sum_{i=1}^n X_i = j - a\right) = \mathbb{P}\left(\sum_{i=1}^n X_i = j + b - a - b\right) = \mathbb{P}(S_n = j + b | S_0 = a + b)$$
2.
$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}\left(\sum_{i=1}^n X_i = j - a\right) = \mathbb{P}\left(\sum_{i=m+1}^{n+m} X_i = j - a\right) = \mathbb{P}(S_{m+n} = j | S_m = a)$$

3. Jeżeli znamy wartość S_m , to rozkład S_{m+n} zależy tylko od wartości X_{m+1}, \dots, X_{m+n} . \square

Odnosząc się do analogii ze wstępu, własność pierwszą można interpretować następująco: pijak wracający z baru chwycie się tak samo, niezależnie od tego gdzie się znajduje. Co więcej (własność druga), będzie on miał taki sam problem z zachowaniem kierunku marszu niezależnie od tego jak dawno opuścił bar (matematyka zajmuje się modelami idealnymi – w związku z tym np. stan upojenia alkoholowego jest permanentny, przez co nie istnieje również kac). Porzuciwszy interpretację barową własność trzecią można skwitować filozoficznym stwierdzeniem „przyszłość zależy od przeszłości tylko poprzez teraźniejszość”.

1.2. Bariery pochłaniające

Wyobraźmy sobie, że idziemy do kasyna z silnym postanowieniem zdobycia pieniędzy na samochód, który kosztuje $\$N$ (ponieważ samochody są drogie N będzie duże). Kasyno opuszczamy z chwilą wygrania założonej kwoty. Oczywiście nic tam po nas, jeśli wszystko przegramy – wtedy nie pozostaje nic innego jak powrót do domu (na piechotę). W takich sytuacjach mówimy o *spacerze losowym z barierami pochłaniającymi* (lewą w 0 i prawą w N).

Przykład 1.4. Gracz zaczyna grę z kapitałem k , $0 \leq k \leq N$ i kończy grę z chwilą wygrania kwoty N lub przegrania wszystkich pieniędzy. Interesuje nas prawdopodobieństwo bankructwa gracza¹.

Niech A_k oznacza zdarzenie, że gracz zaczynający z kapitałem k przegra wszystkie pieniądze, czyli dojdzie do 0 zanim dojdzie do N . Niech $p_k = \mathbb{P}(A_k)$. Sformułowanie problemu podpowiada nam następujące warunki brzegowe: $p_0 = 1$, $p_N = 0$. Wówczas dla $1 \leq k \leq N - 1$ mamy

$$p_k = \mathbb{P}(A_k | X_1 = 1)p + \mathbb{P}(A_k | X_1 = -1)q = p\mathbb{P}(A_{k+1}) + q\mathbb{P}(A_{k-1}) = pp_{k+1} + qp_{k-1}$$

Dla tego równania rekurencyjnego otrzymujemy następujące równanie charakterystyczne:

$$px^2 - x + q = 0,$$

którego pierwiastkami są $x_1 = 1$ i $x_2 = \frac{q}{p}$.

Rozważmy teraz dwa przypadki:

(a) $p \neq q$

Z założenia mamy $x_1 \neq x_2$, zatem

$$p_k = Ax_1^k + Bx_2^k.$$

Z warunków początkowych otrzymujemy układ równań

$$\begin{cases} 1 = A + B \\ 0 = A + B \left(\frac{q}{p}\right)^N \end{cases}$$

Rozwiązanie układu równań daje nam wynik

$$p_k = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}$$

(b) $p = q$

Z założenia mamy $x_1 = x_2 = 1$, zatem

$$p_k = A + Bk.$$

¹ Równie dobrze moglibyśmy pytać o prawdopodobieństwo wygrania pieniędzy na samochód, jednak będziemy przeważnie pytali o prawdopodobieństwo bankructwa. Prawdopodobnie wynika to z zawistnej natury ludzkiej.

Z warunków początkowych otrzymujemy układ równań

$$\begin{cases} 1 = A \\ 0 = A + BN \end{cases}$$

Rozwiązanie układu równań daje nam wynik

$$p_k = 1 - \frac{k}{N}$$

Przykład 1.5. Tym razem pytamy o oczekiwany czas gry z Przykładu 1.4. Niech H_k oznacza liczbę kroków wykonanych do momentu osiągnięcia którejkolwiek z barier pod warunkiem, że zaczynamy w k oraz niech $D_k = \mathbb{E}(H_k)$. Z warunków zadania mamy $D_0 = D_N = 0$. Dla każdego $1 \leq k \leq N - 1$ mamy

$$\begin{aligned} D_k &= \mathbb{E}(H_k | X_1 = 1)p + \mathbb{E}(H_k | X_1 = -1)q = (D_{k+1} + 1)p + (D_{k-1} + 1)q = \\ &= pD_{k+1} + qD_{k-1} + 1 \end{aligned}$$

Niech $p \neq q$. Przypuśćmy, że rozwiązanie jest postaci $D_k = \alpha k$, gdzie α jest pewną stałą. Wówczas

$$\alpha k = p\alpha(k+1) + q\alpha(k-1) + 1 = kp\alpha + p\alpha + kq\alpha - q\alpha + 1 = \alpha k + p\alpha - q\alpha + 1$$

$$p\alpha - q\alpha + 1 = 0$$

$$\alpha = \frac{1}{1-2p}$$

Zatem $D_k = \frac{k}{1-2p}$ jest szczególnym rozwiązaniem równania $D_k = pD_{k+1} + qD_{k-1} + 1$. Na podstawie Przykładu 1.4 wiemy, że rozwiązanie ogólne równania jednorodnego jest postaci $A + B\left(\frac{q}{p}\right)^k$, zatem rozwiązanie naszego równania rekurencyjnego będzie postaci

$$D_k = \frac{k}{1-2p} + A + B\left(\frac{q}{p}\right)^k$$

Z warunków początkowych otrzymujemy

$$\begin{cases} 0 = A + B \\ 0 = \frac{N}{1-2p} + A + B\left(\frac{q}{p}\right)^N, \end{cases}$$

a stąd

$$A = \frac{N}{(1-2p)\left(\left(\frac{q}{p}\right)^N - 1\right)}$$

$$B = \frac{-N}{(1-2p)\left(\left(\frac{q}{p}\right)^N - 1\right)}$$

Zatem

$$\begin{aligned} D_k &= \frac{k}{1-2p} + \frac{N}{(1-2p)\left(\left(\frac{q}{p}\right)^N - 1\right)} - \frac{N}{(1-2p)\left(\left(\frac{q}{p}\right)^N - 1\right)} \left(\frac{q}{p}\right)^k = \\ &= \frac{k}{1-2p} + \frac{N}{(1-2p)\left(\left(\frac{q}{p}\right)^N - 1\right)} \left(1 - \left(\frac{q}{p}\right)^k\right) \end{aligned}$$

Zajmiemy się teraz przypadkiem symetrycznym, $p = q = \frac{1}{2}$. Wówczas $D_k = D_{N-k}$, a równanie rekurencyjne jest postaci

$$D_k - D_{k-1} = D_{k+1} - D_k + 2$$

Oznaczmy $B_k = D_k - D_{k-1}$.

$$B_k = B_{k+1} + 2$$

$$B_1 = D_1$$

$$B_N = -D_{N-1} = -D_1$$

$$B_1 = B_N + 2(N-1) \Rightarrow D_1 = D_{N-1} = N-1$$

$$B_k = B_N + 2(N-k) = N - 2k + 1$$

$$D_2 = B_2 + D_1 = N - 3 + N - 1 = 2(N-2) = D_{N-2}$$

Udowodnimy indukcyjnie, że $D_k = k(N-k)$.

$$D_k = B_k + D_{k-1} = N - 2k + 1 + (k-1)(N-k+1) = k(N-k)$$

1.3. Bariery odpychające

Weźmy teraz pod uwagę następującą sytuację: grający ma bogatego wuja, który, w zależności od nastroju, przesyła z prawdopodobieństwem p siostrzeńcowi marnotrawnemu \$1 gdy ten przegra wszystkie swoje pieniądze. Mówimy wtedy, że dolna bariera jest *odpychająca*.

Przykład 1.6. Niech S_n będzie *PSL* z dolną barierą odpychającą, a górną pochłaniającą, czyli

$$\mathbb{P}(S_{n+1} = 1 | S_n = 0) = p,$$

$$\mathbb{P}(S_{n+1} = 0 | S_n = 0) = q,$$

co oznacza, że przy osiągnięciu 0 z prawdopodobieństwem p przejdziemy do 1, a z prawdopodobieństwem $q = 1-p$ zostaniemy w 0 (nie możemy zejść poniżej 0). Niech G_k oznacza liczbę kroków wykonanych do momentu osiągnięcia górnej bariery pod warunkiem, że zaczynamy w k . Niech $F_k = \mathbb{E}(G_k)$. Równanie rekurencyjne dla tego przykładu jest takie samo jak w Przykładzie 1.5, jednak z innymi warunkami początkowymi: $F_N = 0$, $F_0 = p(F_1 + 1) + q(F_0 + 1)$.

2. Wykład 2: PSL – metoda zliczania ścieżek

Wstęp

Będziemy dalej studiować zachowania osobnika, którego grą zajmowaliśmy się na Wykładzie 1. Tym razem bardziej interesować nas będą jego pełne wrażeń wędrówki w stanie upojenia (np. zwycięstwo w kasynie). Dowiemy się, że jego kochająca (lub nie) żona może mieć spore problemy z odnalezieniem swojego męża jeśli go nie zastanie w kasynie, ponieważ w trakcie wędrówek, pomiędzy przegrywaniem kolejnych dolarów, odwiedza on wszystkich możliwych znajomych i nieznajomych. Ponadto dowiemy się w jaki sposób należy obstawiać zwycięstwo jednego z kandydatów w wyborach w trakcie kolejnego wyczytywania głosów, które otrzymali.

2.1. Zasada odbicia

Niech X_1, X_2, \dots będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie prawdopodobieństwa (*iid*):

$$X_i = \begin{cases} 1 & \text{z prawdopodobieństwem } p \\ -1 & \text{z prawdopodobieństwem } q = 1 - p \end{cases}$$

Rozważmy prosty spacer losowy

$$S_n = S_0 + \sum_{i=1}^n X_i$$

Zakładając, że zaczynamy w punkcie $S_0 = a$ możemy policzyć prawdopodobieństwo, że po n krokach znajdziemy się w punkcie b :

$$\mathbb{P}(S_n = b | S_0 = a) = p^{\frac{n+b-a}{2}} q^{\frac{n-b+a}{2}} \binom{n}{\frac{n+b-a}{2}}$$

Jest tak, gdyż idąc z punktu a do b w n krokach niezależnie od trajektorii zawsze musimy wykonać tyle samo kroków w górę i tyle samo kroków w dół, trzeba jedynie wybrać ich pozycje. Wprowadźmy na chwilę zmienne pomocnicze:

$$r = \frac{n+b-a}{2}, l = \frac{n-b+a}{2}$$

Wówczas to prawdopodobieństwo możemy zapisać następująco:

$$\mathbb{P}(S_n = b | S_0 = a) = \binom{n}{r} p^r q^l = \binom{n}{l} p^r q^l \quad (2.1.1)$$

Wprowadźmy następujące oznaczenia: Liczbę wszystkich trajektorii (ścieżek) od a do b w n krokach będziemy oznaczać przez $N_n(a, b)$. Liczbę wszystkich trajektorii (ścieżek) od a do b w n krokach przecinających oś OX będziemy oznaczać przez $N_n^0(a, b)$.

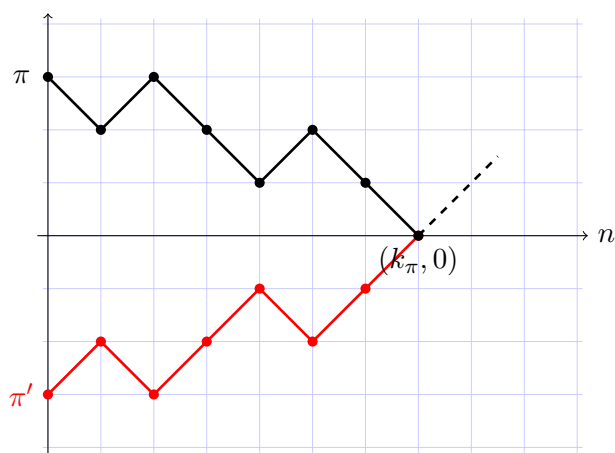
Fakt 2.1:

$$N_n(a, b) = \binom{n}{r} = \binom{n}{l} = \binom{n}{\frac{n+b-a}{2}}$$

Zasada 2.2 (Zasada odbicia). Liczba wszystkich trajektorii od a do b , które przecinają oś OX , jest równa liczbie wszystkich trajektorii od $-a$ do b . Innymi słowy:

$$N_n^0(a, b) = N_n(-a, b)$$

Dowód. Niech π będzie dowolną trajekcją z a do b . Przez $(k_\pi, 0)$ oznaczmy punkt pierwszego kontaktu π z osią OX . Niech π' będzie trajekcją powstałą poprzez odbicie π względem OX na odcinku od 0 do $(k_\pi, 0)$. Wówczas π' jest trajekcją od $-a$ do b , a powyższe przyporządkowanie bijekcją pomiędzy zbiorami $N_n^0(a, b)$ i $N_n(-a, b)$ \square



Rysunek 2.1. Odbijanie ścieżki od $(0, 0)$ do $(k_\pi, 0)$

Wniosek 2.3 (Whitworth 1878 – Ballot Theorem). Niech $b > 0$. Liczba ścieżek z $(0, 0)$ do (n, b) , które na odcinku $\langle (1, 0), (n, 0) \rangle$ nie przecinają osi OX wynosi:

$$\frac{b}{n} \binom{n}{\frac{n+b}{2}}$$

Dowód. Zauważmy, że pierwszy ruch musi być na pozycję $(1, 1)$. Mamy zatem:

$$\begin{aligned}
& N_{n-1}(1, b) - N_{n-1}^0(1, b) \stackrel{2.2}{=} N_{n-1}(1, b) - N_{n-1}(-1, b) \stackrel{2.1}{=} \\
& = \binom{n-1}{\frac{1}{2}(n-1+b-1)} - \binom{n-1}{\frac{1}{2}(n-1+b+1)} = \\
& = \frac{(n-1)!}{(\frac{1}{2}(n-2+b))!(n-1-\frac{1}{2}(n-2+b))!} - \frac{(n-1)!}{(\frac{1}{2}(n+b))!(n-1-\frac{1}{2}(n+b))!} = \\
& = \frac{(n-1)!}{(\frac{1}{2}(n+b)-1)!(n-\frac{1}{2}(n+b))!} - \frac{(n-1)!}{(\frac{1}{2}(n+b))!(n-1-\frac{1}{2}(n+b))!} = \tag{2.1.2} \\
& = \frac{(n-1)!(\frac{1}{2}(n+b))}{(\frac{1}{2}(n+b))!(n-\frac{1}{2}(n+b))!} - \frac{(n-1)!(n-\frac{1}{2}(n+b))}{(\frac{1}{2}(n+b))!(n-\frac{1}{2}(n+b))!} = \\
& = \frac{1}{n} \left(\frac{1}{2}(n+b) - (n - \frac{1}{2}(n+b)) \right) \frac{n!}{(\frac{1}{2}(n+b))!(n-\frac{1}{2}(n+b))!} = \\
& = \frac{b}{n} \binom{n}{\frac{n+b}{2}}
\end{aligned}$$

□

Zadanie 2.4. Mamy dwóch kandydatów startujących w wyborach prezydenckich o imionach Kaczor i Donald. Kaczor otrzymał α głosów, a Donald β , gdzie (co skądinąd wiemy) $\alpha > \beta$. Co więcej cały naród uczestniczył w następującej grze: w trakcie wieczoru wyborczego podczas wyczytywania (i liczenia) głosów oddanych na kandydatów, zakłady bukmacherskie pozwalają obstawiać (np. przez internet) zwycięzcę. Potrzebują jednak algorytmu (wszystko musi odbywać się w czasie rzeczywistym), który pozwoli im ustalić takie stawki, aby na tym zarobić. Na razie wiedzą jedynie jakie jest prawdopodobieństwo, że podczas wyczytywania głosów Kaczor będzie prowadził:

$$P = \frac{\#\{\text{Wszystkie ścieżki od } (0, 0) \text{ do } (\alpha + \beta, \alpha - \beta) \text{ nie przecinające osi } OX\}}{\#\{\text{Wszystkie ścieżki od } (0, 0) \text{ do } (\alpha + \beta, \alpha - \beta)\}} = \frac{\alpha - \beta}{\alpha + \beta}$$

2.2. Powroty do osi

Można sobie postawić pytanie: jakie jest prawdopodobieństwo, że spacer poza kasynem będzie miał długość przynajmniej n ?

Wniosek 2.5. Jeśli $S_0 = 0$, to dla każdego $n \geq 1$ mamy:

$$\mathbb{P}(S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = b) = \frac{|b|}{n} \mathbb{P}(S_n = b)$$

A stąd wynika, że również:

$$\mathbb{P}(S_1 \neq 0, \dots, S_n \neq 0) = \frac{1}{n} \mathbb{E}(|S_n|)$$

Dowód. Niech $b > 0$. Wtedy

$$\mathbb{P}(S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = b) \stackrel{2.3}{=} \frac{b}{n} N_n(0, b) p^{\frac{n+b}{2}} q^{\frac{n-b}{2}} \stackrel{2.1}{=} \frac{b}{n} \mathbb{P}(S_n = b)$$

Analogicznie możemy stwierdzić, że dla $b < 0$ zachodzi

$$\mathbb{P}(S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = b) = \frac{-b}{n} N_n(0, -b) p^{\frac{n-b}{2}} q^{\frac{n+b}{2}} = \frac{-b}{n} \mathbb{P}(S_n = b)$$

zatem ogólny wzór zapisujemy jako

$$\mathbb{P}(S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = b) = \frac{|b|}{n} \mathbb{P}(S_n = b)$$

Czyli pierwsza równość została udowodniona. Drugą można pokazać przez następujące przejścia:

$$\begin{aligned} \mathbb{P}(S_1 \neq 0, \dots, S_n \neq 0) &= \sum_{b \neq 0} \mathbb{P}(S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = b) = \\ &= \sum_b \frac{|b|}{n} \mathbb{P}(S_n = b) = \frac{1}{n} \mathbb{E}(|S_n|) \end{aligned}$$

□

Oznaczmy przez M_n maksymalną wartość jaką osiągnie S_i na odcinku do 0 do n :

$$M_n = \max_{0 \leq i \leq n} S_i.$$

Dla $S_0 = 0$ mamy oczywiście $M_n \geq 0$.

Twierdzenie 2.6.

Dla każdego $r \geq 1$ mamy:

$$\mathbb{P}(M_n \geq r, S_n = b) = \begin{cases} \mathbb{P}(S_n = b) & b \geq r \\ \left(\frac{q}{p}\right)^{r-b} \mathbb{P}(S_n = 2r - b) & b < r \end{cases}$$

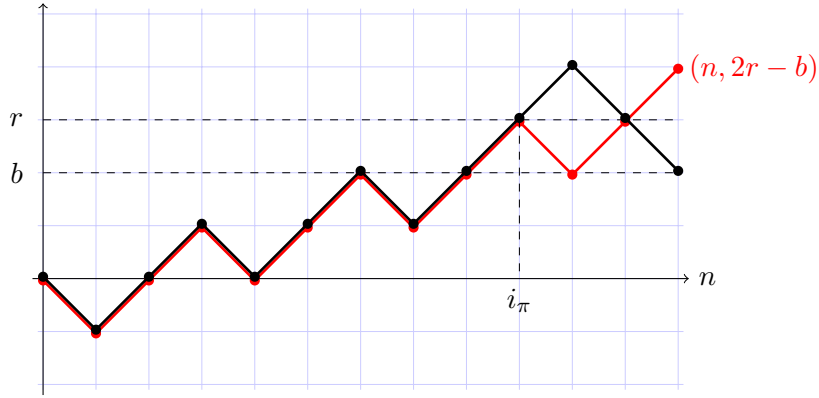
Wniosek 2.7. Dla $p = q$ mamy:

$$\mathbb{P}(M_n \geq r) = \mathbb{P}(S_n \geq r) + \mathbb{P}(S_n \geq r + 1)$$

Dowód.

$$\begin{aligned} \mathbb{P}(M_n \geq r) &= \sum_b \mathbb{P}(M_n \geq r, S_n = b) \stackrel{2.6}{=} \sum_{b=r}^n \mathbb{P}(S_n = b) + \sum_{b=-\infty}^{r-1} \left(\frac{q}{p}\right)^{r-b} \mathbb{P}(S_n = 2r - b) = \\ &= \mathbb{P}(S_n \geq r) + \sum_{c=r+1}^{\infty} \left(\frac{q}{p}\right)^{c-r} \mathbb{P}(S_n = c) = \mathbb{P}(S_n \geq r) + \mathbb{P}(S_n \geq r + 1) \end{aligned}$$

□



Rysunek 2.2. Odbicie symetryczne ścieżki względem prostej $y = r$

Dowód Twierdzenia 2.6. Bez straty ogólności możemy przyjąć, że $r \geq 1$ i $b < r$. Niech $N_n^r(0, b)$ oznacza liczbę ścieżek π z $(0, 0)$ do (n, b) , które przecinają prostą $y = r$. Ponadto niech $i_\pi = \min_{(i,r) \in \pi} \{i\}$ oznacza pierwszy moment czasu, w którym osiągnięta zostanie prosta $y = r$. Możemy symetrycznie odbić względem prostej $y = r$ segment ścieżki dla $i_\pi \leq x \leq n$. Tak określone przekształcenie wyznacza bijekcję między zbiorami.

$$N_n^r(0, b) = N_n(0, 2r - b)$$

Wówczas:

$$\begin{aligned} \mathbb{P}(M_n \geq r, S_n = b) &\stackrel{2.1}{=} N_n^r(0, b) p^{\frac{n+b}{2}} q^{\frac{n-b}{2}} = N_n(0, 2r - b) p^{\left(\frac{n+2r-b}{2} + \frac{2b-2r}{2}\right)} q^{\left(\frac{n-2r+b}{2} + \frac{2r-2b}{2}\right)} \stackrel{2.1}{=} \\ &= \left(\frac{q}{p}\right)^{r-b} \mathbb{P}(S_n = 2r - b) \end{aligned}$$

□

2.3. Zagadnienie pierwszych odwiedzin

W niniejszym rozdziale opisane zostanie prawdopodobieństwo zdarzenia, że startując z $(0, 0)$ osiągniemy punkt $b > 0$ po raz pierwszy w n -tym kroku.

Definicja 2.8. Niech $f_b(n)$ oznacza prawdopodobieństwo, że startując z 0 trafimy w punkt b po raz pierwszy w n -tym kroku ($n \geq 1$).

$$f_b(n) = \begin{cases} \mathbb{P}(S_0 = 0, S_1, \dots, S_{n-1} < b, S_n = b) & b > 0 \\ \mathbb{P}(S_0 = 0, S_1, \dots, S_{n-1} > b, S_n = b) & b < 0 \end{cases}$$

Twierdzenie 2.9 (Hitting time Theorem).

$$f_b(n) = \frac{|b|}{n} \mathbb{P}(S_n = b)$$

Dowód. Ograniczymy się do dowodu dla $b > 0$. Dla $b < 0$ dowód przebiega analogicznie.

$$\begin{aligned}
f_b(n) &\stackrel{2.8}{=} \mathbb{P}(M_{n-1} = b-1, S_{n-1} = b-1, S_n = b) = \\
&= \mathbb{P}(S_n = b | M_{n-1} = b-1, S_{n-1} = b-1) \mathbb{P}(M_{n-1} = b-1, S_{n-1} = b-1) = \\
&= p \mathbb{P}(M_{n-1} = b-1, S_{n-1} = b-1) = \\
&= p [\mathbb{P}(M_{n-1} \geq b-1, S_{n-1} = b-1) - \mathbb{P}(M_{n-1} \geq b, S_{n-1} = b-1)] = \\
&\stackrel{2.6}{=} p \left[\mathbb{P}(M_{n-1} \geq b-1 | S_{n-1} = b-1) \mathbb{P}(S_{n-1} = b-1) - \left(\frac{q}{p}\right)^{b-(b-1)} \mathbb{P}(S_{n-1} = 2b - (b-1)) \right] = \\
&= p \left[\mathbb{P}(S_{n-1} = b-1) - \left(\frac{q}{p}\right) \mathbb{P}(S_{n-1} = b+1) \right] = \\
&\stackrel{2.1.1}{=} p \binom{n-1}{\frac{1}{2}(n+b-2)} p^{\frac{1}{2}(n+b-2)} q^{\frac{1}{2}(n-b)} - q \binom{n-1}{\frac{1}{2}(n+b)} p^{\frac{1}{2}(n+b)} q^{\frac{1}{2}(n-b-2)} = \\
&= \left(\binom{n-1}{\frac{1}{2}(n+b-2)} - \binom{n-1}{\frac{1}{2}(n+b)} \right) p^{\frac{1}{2}(n+b)} q^{\frac{1}{2}(n-b)} = \\
&\stackrel{2.1}{=} \frac{b}{n} N_n(0, b) p^{\frac{1}{2}(n+b)} q^{\frac{1}{2}(n-b)} = \\
&\stackrel{2.1.1}{=} \frac{b}{n} \mathbb{P}(S_n = b)
\end{aligned}$$

□

2.3.1. Technika odwracania

Twierdzenie 2.9 jest w swojej tezie bardzo podobne do Wniosku 2.5. Podobieństwo to nie jest przypadkowe i poniżej zostanie przedstawiona technika (tzw. technika odwracania spaceru losowego), która pokaże, że w istocie oba te twierdzenia są dualne.

Definicja 2.10 (Odwracanie spaceru losowego). Dla niezależnych zmiennych losowych X_1, X_2, \dots o jednakowym rozkładzie (*iid*) *odwróceniem spaceru losowego* (składającego się z n kroków):

$$\{0, S_1, S_2, \dots, S_n\} = \left\{ 0, X_1, X_1 + X_2, \dots, \sum_{i=1}^n X_i \right\}$$

nazywamy następujący spacer (n -krokowy):

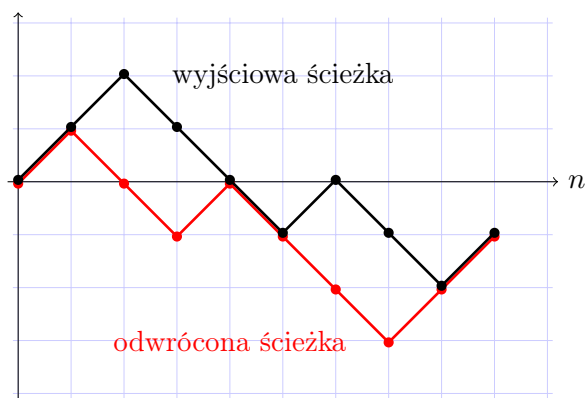
$$\{0, T_1, T_2, \dots, T_n\} = \left\{ 0, X_n, X_n + X_{n-1}, \dots, \sum_{i=1}^n X_i \right\}$$

Zauważmy, że na podstawie powyższej definicji możemy przeformułować jedno twierdzenie w drugie. Czyli dla $n-1 \geq i \geq 1$:

$$S_1, \dots, S_n \neq 0, S_n = b \iff T_n = b, T_n - T_{n-i} = X_1 + \dots + X_i > 0$$

To oznacza, że pierwsza wizyta w b dla odwróconego spaceru nastąpi w chwili n , zatem:

$$\mathbb{P}(S_1, \dots, S_n \neq 0, S_n = b) = f_b(n)$$



Rysunek 2.3. Technika odwracania

Zasadę tę można wykorzystać do obliczenia średniej liczby wizyt w b do pierwszego powrotu do 0.

$$\begin{aligned}
 \mu_b &= \mathbb{E}(\#\text{wizyt w } b \text{ przed pierwszym powrotem do } 0) = \\
 &= \mathbb{E}\left(\sum_{n=1}^{\infty} I_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(I_n) = \sum_{n=1}^{\infty} 1 \cdot \mathbb{P}(I_n = 1) = \sum_{n=1}^{\infty} \mathbb{P}(S_1, \dots, S_n \neq 0, S_n = b) = \\
 &= \sum_{n=1}^{\infty} f_b(n) \stackrel{*}{=} \mathbb{P}(\exists n : S_n = b)
 \end{aligned}$$

gdzie I_n jest zmienną indykatorową:

$$I_n = \begin{cases} 1 & \text{gdy } S_1 \neq 0, \dots, S_n \neq 0, S_n = b \\ 0 & \text{w pozostałych przypadkach} \end{cases}$$

a równość \star jest prawdziwa, gdyż zdarzenia określone w prawdopodobieństwie $f_b(n)$ są rozłączne.

Powyższe uwagi prowadzą do następującego twierdzenia:

Twierdzenie 2.11.

Jeśli $p = q$, to dla dowolnego $b > 0$ zachodzi $\mu_b = 1$.

Powyższe rozważania można zinterpretować w następujący sposób: Rzucamy monetą aż liczba orłów zrówna się z liczbą reszek. Gracz dostaje wypłatę za każdym razem, gdy różnica między liczbą orłów i reszek wynosi b . Zgodnie z twierdzeniem, aby gra była sprawiedliwa, opłata za nią musi wynosić 1 niezależnie od tego ile wynosi b .

3. Wykład 3: Funkcje tworzące

Dokończenie poprzedniego wykładu

Założmy, że $S_0 = 0$, $p = \frac{1}{2}$. Niech T będzie spacerem odwrotnym do S , a T_{2n} oznacza czas ostatniej wizyty w punkcie 0 w ciągu pierwszych $2n$ kroków. Ponadto oznaczmy $\alpha_{2n}(2k) = \mathbb{P}(T_{2n} = 2k)$ oraz $u_{2k} = \mathbb{P}(S_{2k} = 0)$.

Twierdzenie 3.1 (Prawo arcusa sinusa).

$$\alpha_{2n}(2k) = u_{2k} \cdot u_{2n-2k}$$

Przed rozpoczęciem dowodu zauważmy, że

$$u_{2k} = \mathbb{P}(S_{2k} = 0) = \binom{2k}{k} \left(\frac{1}{2}\right)^{2k} \sim \frac{\sqrt{4\pi k} \cdot (2k)^{2k} \cdot e^{-2k}}{(\sqrt{2\pi k} \cdot k^k \cdot e^{-k})^2 \cdot 2^{2k}} = \frac{1}{\sqrt{\pi k}},$$

zatem dla dużych k i $n - k$ mamy

$$\alpha_{2n}(2k) \sim \frac{1}{\pi \sqrt{k(n-k)}}.$$

Ponadto zauważmy jak zachowuje się następująca dystrybuanta:

$$\begin{aligned} F_{\frac{T_{2n}}{2n}}(x) &= \mathbb{P}\left(\frac{T_{2n}}{2n} \leq x\right) = \mathbb{P}(T_{2n} \leq 2nx) \\ &\sim \sum_{k \leq xn} \frac{1}{\pi \sqrt{k(n-k)}} \\ &\sim \int_0^{xn} \frac{dk}{\pi \sqrt{k(n-k)}} \\ &= \frac{2}{\pi} \arcsin(\sqrt{x}). \end{aligned}$$

Jeśli przeformułujemy problem na język orłów i reszek, to mogą zadziwić nas następujące „zaskoczenia”:

- 1) Intuicja podpowiada, że ostatnie zrównanie się liczby orłów i liczby reszek powinno mieć miejsce pod koniec eksperymentu, ale $\alpha_{2n}(2k) = \alpha_{2n}(2n - 2k)$. Dla przykładu mamy:

$$\mathbb{P}\left(T_{2n} \leq \frac{2n}{10}\right) = \mathbb{P}\left(T_{2n} \geq \frac{9}{10}(2n)\right)$$

oraz

$$\mathbb{P}(T_{2n} \leq n) = \mathbb{P}(T_{2n} \geq n)$$

2) Intuicyjnym jest również, że powinno być dużo zrównań liczby orłów i reszek podczas eksperymentu, ale

$$\mathbb{P}(T_{2n} \leq n) = \frac{1}{2}, \quad \mathbb{P}\left(T_{2n} \leq \frac{2n}{10}\right) \sim \frac{2}{\pi} \arcsin\left(\sqrt{\frac{1}{10}}\right) \approx \frac{1}{5}.$$

Dowód twierdzenia 3.1.

$$\begin{aligned} \alpha_{2n}(2k) &= \mathbb{P}(T_{2n} = 2k) = \mathbb{P}(S_{2k} = 0, S_{2k+1} \cdot \dots \cdot S_{2n} \neq 0) = \\ &= \mathbb{P}(S_{2k} = 0) \cdot \mathbb{P}(S_1 \cdot \dots \cdot S_{2n-2k} \neq 0) = \\ &\stackrel{*}{=} \mathbb{P}(S_{2k} = 0) \cdot \mathbb{P}(S_{2n-2k} = 0) = \\ &= u_{2k} \cdot u_{2n-2k} \end{aligned}$$

gdzie pierwsza równość wynika z techniki odwracania spacerów losowych, a równość * wynika z zadania domowego 2.9. \square

3.1. Funkcje tworzące

Definicja 3.2. Funkcję

$$G_{\mathbf{a}}(s) = \sum_{i=0}^{\infty} a_i s^i$$

nazywamy *funkcją tworzącą* dla danego ciągu $\mathbf{a} = (a_0, a_1, \dots)$. Funkcję

$$E_{\mathbf{a}}(s) = \sum_{i=0}^{\infty} \frac{a_i s^i}{i!}$$

nazywamy *wykładniczą funkcją tworzącą* dla danego ciągu \mathbf{a} .

Definicja 3.3. *Splotem (konwolucją)* danych dwóch ciągów \mathbf{a} i \mathbf{b} nazywamy taki ciąg $\mathbf{c} = \mathbf{a} * \mathbf{b}$, że

$$c_n = \sum_{i=0}^n a_i b_{n-i}$$

Ponadto jest jasne, że $G_{\mathbf{c}} = G_{\mathbf{a}} \cdot G_{\mathbf{b}}$.

Przykład 3.4. Niech dane będą takie dwa skończone ciągi \mathbf{a} i \mathbf{b} , że $a_i = b_i = \binom{n}{i} = a_{n-i}$, dla $i = 0, 1, \dots, n$. Niech $\mathbf{c} = \mathbf{a} * \mathbf{a}$, a zatem $c_k = \sum_{i=0}^k \binom{n}{i}^2$ dla $k = 0, 1, \dots, n$. Mamy wtedy

$$G_{\mathbf{c}}(s) = (G_{\mathbf{a}}(s))^2 = ((1+s)^n)^2 = (1+s)^{2n} = \sum_{i=0}^n \binom{2n}{i} s^i,$$

czyli z definicji 3.3

$$c_k = \binom{2n}{k} = \sum_{i=0}^k \binom{n}{i} \binom{n}{k-i},$$

dla każdego $k = 0, 1, \dots, n$.

Definicja 3.5. Niech $X \in \mathbb{Z}$ będzie zmienną losową. Wówczas funkcję

$$G_X(s) = \mathbb{E}(s^X) = \sum_i s^i \mathbb{P}(X = i) = \sum_i s^i p_i = G_{\mathbf{p}}(s)$$

nazywamy *funkcją tworzącą prawdopodobieństwa* zmiennej losowej X .

Przykład 3.6. Dla rozkładu Poissona:

$$G(s) = e^{\lambda(s-1)}$$

Przykład 3.7. Dla rozkładu geometrycznego

$$G(s) = \frac{ps}{1 - s(1 - p)}$$

3.1.1. Liczenie momentów

Okazuje się, że funkcje tworzące są bardzo przydatne w liczeniu kolejnych momentów zmiennych losowych dyskretnych.

Definicja 3.8. *Spadającym iloczynem* nazywać będziemy wartość

$$(a)_k = \frac{a!}{(a - k)!}$$

Definicja 3.9. Wartość $\mathbb{E}((X)_k)$ będziemy nazywali *k-tym momentem silniowym* zmiennej losowej X .

W terminach momentów silniowych można z łatwością opisać wariancję:

$$\text{Var}(X) = \mathbb{E}((X)_2) + \mathbb{E}(X) - (\mathbb{E}(X))^2$$

Twierdzenie 3.10.

Niech X będzie dyskretną zmienną losową o funkcji tworzącej $G(s)$. Wtedy

$$\mathbb{E}((X)_k) = G^{(k)}(1),$$

gdzie

$$G^{(k)}(1) = \lim_{s \rightarrow 1} G^{(k)}(s),$$

gdzie promień zbieżności szeregu jest równy 1.

Dowód. Niech $s < 1$. Po obliczeniu k -tej pochodnej mamy:

$$G^{(k)}(s) = \sum_i s^{i-k} (i)_k p_i = \mathbb{E}(s^{X-k} (X)_k)$$

Dla $s \rightarrow 1^-$ z twierdzenia Abela dla szeregów nieskończonych mamy:

$$G^{(k)}(1) = \lim_{s \rightarrow 1} G^{(k)}(s) = \mathbb{E}((X)_k).$$

□

3.1.2. Sumy zmiennych niezależnych

Funkcje tworzące ułatwiają także określanie rozkładów sum niezależnych zmiennych losowych. Jest tak dzięki prostej zależności opisanej w poniższym twierdzeniu.

Twierdzenie 3.11.

Niech X, Y będą niezależnymi dyskretnymi zmiennymi losowymi. Wtedy

$$G_{X+Y}(s) = G_X(s)G_Y(s)$$

Dowód. Z niezależności X i Y zmienne losowe s^X i s^Y są również niezależne, więc

$$\mathbb{E}(s^X s^Y) = \mathbb{E}(s^X) \mathbb{E}(s^Y)$$

□

Przykład 3.12. Niech S_n będzie zmienną losową o rozkładzie dwumianowym $Bin(n, p)$. Wtedy oczywiście $S_n = \sum_{i=1}^n X_i$, gdzie X_i są niezależnymi zmiennymi losowymi zero–jedynkowymi przyjmującymi wartość 1 z prawdopodobieństwem p i wartość 0 z prawdopodobieństwem $q = 1 - p$. Czyli dla dowolnego i mamy $G_{X_i} = s^0 q + s^1 p = q + ps$. Ostatecznie, korzystając z twierdzenia 3.11:

$$G_{S_n}(s) = (G_{X_1}(s))^n = (q + ps)^n.$$

A co jeśli zmienna losowa jest sumą losowej liczby niezależnych zmiennych losowych?

Twierdzenie 3.13.

Niech X_1, X_2, \dots będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie (*iid*). Oznaczmy $G_{X_i} = G$ dla $i = 1, 2, \dots$. Niech $N \in \{0, 1, 2, \dots\}$ będzie zmienną losową niezależną od X_1, X_2, \dots oraz $S = X_1 + X_2 + \dots + X_N$. Przy tych oznaczeniach zachodzi równość

$$G_S(s) = G_N(G(s)).$$

Dowód. $G_S(s) = \mathbb{E}(s^S) = \mathbb{E}(\mathbb{E}(s^S|N))$. Czyli z niezależności zmiennych N i X_1, X_2, \dots oraz definicji wartości oczekiwanej:

$$\begin{aligned} G_S(s) &= \sum_n \mathbb{E}(s^S|N = n) \mathbb{P}(N = n) = \\ &= \sum_n \mathbb{E}(s^{X_1+X_2+\dots+X_n}) \mathbb{P}(N = n) = \\ &= \sum_n \mathbb{E}(s^{X_1}) \dots \mathbb{E}(s^{X_n}) \mathbb{P}(N = n) = \\ &= \sum_n (G(s))^n \mathbb{P}(N = n) = \\ &\stackrel{3.5}{=} G_N(G(s)). \end{aligned}$$

□

Przykład 3.14. Kura znosi N jaj, gdzie N jest zmienną losową o rozkładzie Poissona $Po(\lambda)$. Z każdego z jaj z prawdopodobieństwem p wykluwa się kurczak, natomiast z prawdopodobieństwem $q = 1 - p$ nie wykluwa się nic. Można zadać pytanie: „Ile wykluje się kurcząt?”. Będzie to pytanie o rozkład zmiennej losowej $S = \sum_{i=1}^N X_i$. Skoro:

$$G_N(s) = \sum_{i=0}^{\infty} \frac{e^{-\lambda} \lambda^i}{i!} s^i = e^{\lambda(s-1)}$$

oraz

$$G_{X_i}(s) = G(s) = q + ps$$

więc z twierdzenia 3.13

$$G_S(s) = G_N(G(s)) = e^{\lambda(q+ps-1)} = e^{\lambda p(s-1)}.$$

Czyli ostatecznie S ma rozkład Poissona $Po(\lambda p)$.

3.1.3. Zastosowanie dla spacerów losowych

„Po obfitej jajecznicy możemy wyjść na spacer”. W spacerze losowym o punkcie początkowym w zerze ($S_0 = 0$) zdefiniujemy czas, po którym po raz pierwszy wrócimy do punktu wyjścia, czyli: $T_0 = \min\{S_n = 0 : n \geq 1\}$. Oczywiście $T_0 \in \{2, 4, \dots\} \cup \{\infty\}$. Chcemy znaleźć rozkład tej zmiennej losowej. Wprowadźmy następujące oznaczenia:

$$\begin{aligned} f_0(0) &= 0 \\ f_0(n) &= \mathbb{P}(T_0 = n) = \mathbb{P}(S_1, \dots, S_{n-1} \neq 0, S_n = 0), \text{ dla } n \geq 1 \\ p_0(n) &= \mathbb{P}(S_n = 0) \end{aligned}$$

Przy tych oznaczeniach zmienne losowe T_0 i S_n mają funkcje tworzące:

$$\begin{aligned} P_0(s) &= \sum_{n=0}^{\infty} p_0(n) s^n \\ F_0(s) &= \sum_{n=0}^{\infty} f_0(n) s^n \end{aligned}$$

Twierdzenie 3.15.

Następujące równości są prawdziwe:

- (a) $P_0 = 1 + P_0 F_0$
- (b) $P_0(s) = \frac{1}{\sqrt{1-4pqs^2}}$
- (c) $F_0(s) = 1 - \sqrt{1-4pqs^2}$

Dowód. Udowodnijmy najpierw punkt (a) twierdzenia. Zauważmy, że dla dowolnego $n \geq 1$

$$p_0(n) = \sum_{k=1}^n f_0(k) p_0(n-k),$$

gdyż jeśli w k -tej chwili jesteśmy w punkcie wyjściowym, to tak, jak byśmy zaczęli spacer od nowa. Czyli, mówiąc nieformalnie, sumujemy po wszystkich momentach bycia wcześniej w zerze po raz pierwszy, a potem „resetujemy” czas.

Korzystając z operacji splotu, funkcji tworzących P_0 i F_0 oraz pamiętając, że $p_0(0) = 1$ i $f_0(0) = 0$ mamy:

$$\begin{aligned} P_0(s) - 1 &= \sum_{n=0}^{\infty} p_0(n)s^n - 1 = \sum_{n=1}^{\infty} p_0(n)s^n = \\ &= \sum_{n=1}^{\infty} \left(\sum_{k=1}^n f_0(k)p_0(n-k) \right) s^n = \\ &= \left(\sum_{m_1=1}^{\infty} f_0(m_1)s^{m_1} \right) \left(\sum_{m_2=0}^{\infty} p_0(m_2)s^{m_2} \right) = F_0(s)P_0(s). \end{aligned}$$

Punkt (b) twierdzenia wynika z tego, iż

$$p_0(n) = \binom{n}{n/2} (pq)^{\frac{n}{2}}.$$

Ostatecznie, korzystając z rozwiązania zadania domowego 3.6 mamy:

$$p_0(s) = \frac{1}{\sqrt{1-4pqs^2}}$$

Punkt (c), po prostych przeliczeniach, wynika bezpośrednio z (a) i (b). □

Wniosek 3.16.

(i) Jeśli $p = q = \frac{1}{2}$ to z prawdopodobieństwem 1 wrócimy do punktu wyjścia gdyż:

$$\sum_{n=1}^{\infty} f_0(n) = F_0(1) = 1 - \sqrt{1-4pq} = 1 - \sqrt{(p-q)^2} = 1 - |p-q|$$

(ii) Dla $p = \frac{1}{2}$ wartość oczekiwana czasu powrotu jest nieskończona, gdyż skoro:

$$\lim_{s \rightarrow 1^-} F'_0(s) = \lim_{s \rightarrow 1^-} \left(1 - \sqrt{1-4pqs^2} \right)' = \lim_{s \rightarrow 1^-} \frac{2s}{2\sqrt{1-s^2}} = \infty,$$

więc $\mathbb{E}(T_0) = F'_0(1) = \infty$.

4. Wykład 4: Funkcje tworzące. Procesy gałązkowe

Wprowadzenie

Kontynuując nasz spacer spróbujemy w sposób matematyczny spojrzeć na zjawisko *déjà vu*, wstąpimy do kasyna, by oddać się po raz kolejny dobrze nam znanej, lecz wciąż niezmiernie fascynującej grze w orła i reszkę, a kiedy w końcu wreszcie wyjdziemy na zero przyjrzymy się bliżej niezwyklej populacji monet.

4.1. Funkcje tworzące (dokończenie)

4.1.1. Moment pierwszych odwiedzin

Bywa tak, że wstępując do lokalu odnosimy wrażenie, że miejsce to jest nam dziwnie znajome – znajome stoliki, znajome wnętrza, znajomy barman... – zupełnie tak, jakbyśmy kiedyś już tu byli. Zaczynamy się wtedy zastanawiać: kiedy to było? Jeśli nie jest to tylko złudzenie (tzw. *déjà vu*), to w znalezieniu odpowiedzi na to pytanie może nam pomóc aparat funkcji tworzących.

Niech

$$f_r(n) = \mathbb{P}(S_1 \neq r, \dots, S_{n-1} \neq r, S_n = r)$$

oznacza prawdopodobieństwo, że punkt r osiągniemy po raz pierwszy w n -tym kroku. Funkcję tworzącą tego ciągu oznaczmy przez

$$F_r(s) = \sum_{n=1}^{\infty} f_r(n)s^n.$$

Zwróćmy uwagę, że może zachodzić $F_r(1) < 1$, ponieważ może się zdarzyć, że nigdy nie osiągniemy punktu r (istnieją spacery, które nie docierają do r).

Twierdzenie 4.1.

Zachodzą następujące równości:

- (a) $F_r(s) = (F_1(s))^r$ dla $r \geq 1$.
- (b) $F_1(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2qs}$.

Intuicyjnie część (a) tego twierdzenia należy rozumieć w ten sposób, że aby dojść do punktu r musimy wprawdzie pokonać dystans jednego kroku w prawo, potem kolejny taki sam dystans i tak dalej („krok po kroku” albo inaczej „ziarnko do ziarnka aż zbierze się miarka”).

Dowód twierdzenia 4.1.

- (a) Znajdźmy prawdopodobieństwo, że w n -tym kroku po raz pierwszy pojawimy się w punkcie r . Dla $r = 1$ teza jest oczywista. W dalszej części dowodu będziemy zakładać, że $r > 1$. Korzystając ze wzoru na prawdopodobieństwo całkowite otrzymujemy:

$$f_r(n) = \sum_{k=1}^{n-1} f_1(k) f_{r-1}(n-k)$$

(aby dojść z zera do r , musimy najpierw dojść do punktu 1 w k krokach, a następnie w $n-k$ krokach z punktu 1 do punktu r). Po pomnożeniu obustronnie przez s^n i zsumowaniu po n dostajemy

$$\sum_{n=1}^{\infty} f_r(n) s^n = \sum_{n=1}^{\infty} \sum_{k=1}^{n-1} f_1(k) f_{r-1}(n-k) s^n,$$

co z kolei ze wzoru na mnożenie szeregów możemy dalej przekształcić na

$$\sum_{n=1}^{\infty} f_r(n) s^n = \left(\sum_{i=1}^{\infty} f_{r-1}(i) s^i \right) \left(\sum_{j=1}^{\infty} f_1(j) s^j \right),$$

czyli

$$F_r(s) = F_{r-1}(s) F_1(s).$$

Postępując analogicznie dla $F_{r-1}(s)$, $F_{r-2}(s)$... otrzymujemy tezę.

- (b) Niech zmienna losowa $T_r = \min\{n: S_n = r\}$ oznacza moment pierwszej wizyty w punkcie r . Prawdopodobieństwo, że do 1 dojdziemy w pierwszym kroku, jest oczywiście równe $\mathbb{P}(T_1 = 1) = f_1(1) = p$. Niech teraz $n > 1$. Ze wzoru na prawdopodobieństwo całkowite

$$\mathbb{P}(T_1 = n) = \underbrace{\mathbb{P}(T_1 = n | X_1 = 1)}_{=0} p + \mathbb{P}(T_1 = n | X_1 = -1) q.$$

Prawdopodobieństwo $\mathbb{P}(T_1 = n | X_1 = 1)$ jest zerowe, gdyż, jeśli zrobiliśmy pierwszy krok w prawo, to znaczy, że przybyliśmy do punktu 1 „za wcześnie”. Z kolei z jednorodności czasowej i przestrzennej (zobacz Twierdzenie 1.3) widać, że

$$\mathbb{P}(T_1 = n | X_1 = -1) = \mathbb{P}(T_1 = n - 1 | S_0 = -1) = \mathbb{P}(T_2 = n - 1),$$

a stąd

$$\begin{aligned} \mathbb{P}(T_1 = n) &= \mathbb{P}(T_2 = n - 1) q, \\ f_1(n) &= f_2(n - 1) q. \end{aligned}$$

Po obustronnym pomnożeniu przez s^n i zsumowaniu po n dostajemy

$$\begin{aligned} \sum_{n=2}^{\infty} f_1(n)s^n &= q \sum_{n=2}^{\infty} f_2(n-1)s^n \\ \sum_{n=1}^{\infty} f_1(n)s^n - f_1(1)s &= sq \sum_{n=2}^{\infty} f_2(n-1)s^{n-1} \\ \sum_{n=1}^{\infty} f_1(n)s^n - ps &= sq \sum_{n=1}^{\infty} f_2(n)s^n \\ F_1(s) - ps &= sqF_2(s) \\ F_1(s) - ps &= qs(F_1(s))^2 \\ qs(F_1(s))^2 - F_1(s) + ps &= 0 \end{aligned}$$

Powyzsze równanie kwadratowe ma pierwiastki

$$F_1(s)_1 = \frac{1 - \sqrt{1 - 4pqs^2}}{2qs}, \quad F_1(s)_2 = \frac{1 + \sqrt{1 - 4pqs^2}}{2qs}.$$

Ponieważ wiemy, że $F_1(0) = 0$, a z drugiej strony

$$F_1(0)_2 = \lim_{s \rightarrow 0} \frac{1 + \sqrt{1 - 4pqs^2}}{2qs} = \infty,$$

zatem należy odrzucić rozwiązanie $F_1(0)_2$. Zostaje

$$F_1(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2qs}.$$

□

Wniosek 4.2. Prawdopodobieństwo, że spacer kiedykolwiek wejdzie na dodatnią stronę osi, wynosi

$$F_1(1) = \frac{1 - |p - q|}{2q} = \min \left\{ 1, \frac{p}{q} \right\}.$$

Przypomnijmy Twierdzenie 2.9, które mówi, że

$$f_1(n) = \frac{1}{n}P(S_n = 1).$$

Podstawiając ten wynik do definicji funkcji tworzącej F_r otrzymujemy

$$F_1(1) = \sum_{n=1}^{\infty} f_1(n) = \sum_{n=1}^{\infty} \frac{1}{n}P(S_n = 1).$$

Podejrzliwy umysł zastanawia się, czy otrzymane wyniki są sobie równe...

Czytelnikowi zostawiamy dowód tego faktu (jako zadanie domowe 4.3) oraz wskazówkę, że dla $p = q$ i $n = 2m - 1$:

$$F_1(1) = \sum_{n=1}^{\infty} \frac{1}{n} \binom{n}{\frac{n+1}{2}} \frac{1}{2^n} = \sum_{m=1}^{\infty} \frac{(2m-2)!}{m!(m-1)!} \frac{2}{2^{2m}}.$$

Wystarczy pokazać, że powyższa suma równa jest wyrażeniu

$$\sum_{m=1}^{\infty} (-1)^{m+1} \binom{\frac{1}{2}}{m} = 1 - \sqrt{1-1} = 1.$$

4.1.2. Czas przebywania na dodatniej półosi

Wstąpmy jeszcze na chwilę do kasyna. Tutaj krupier proponuje nam ciekawą grę. Rzucamy $2n$ razy pod rząd monetą. Ilekroć wyrzucimy orła, dostajemy dolara, zaś kiedy wyrzucimy reszkę, musimy jednego dolara oddać. Po skończonej grze wyszliśmy na zero – wyrzuciliśmy tyle samo orłów, ile reszek. Mina nam trochę zrzęda – liczyliśmy, że zyskamy choć dolara. Myśl ucieka nam do tych radosnych chwil, kiedy skończona przed chwilą gra jeszcze trwała, a my cieszyliśmy się przewagą naszych dzielnych orłów nad hordami podłych reszek. . . Jak długo trwały te szczęśliwe chwile? Jaki czas utrzymywaliśmy się „nad kreską”?

Doświadczenie to można traktować jako prosty spacer losowy o długości $2n$, który rozpoczął i zakończył się w zerze. Będziemy rozważać warunkową przestrzeń probabilistyczną – pod warunkiem, że $S_{2n} = 0$. Niech zmienna losowa L_{2n} oznacza liczbę kroków pobytu na dodatniej półosi w czasie $\{1, 2, \dots, 2n\}$. Za krok będziemy tym razem uważali przejście z jednego punktu do drugiego, przy czym, żeby uznać krok za dodatni, wystarczy, że dodatni będzie jego punkt początkowy bądź końcowy. Innymi słowy, dodatnimi krokami są te odcinki wykresu, które znajdują się nad osią odciętych.

Zauważmy, że rozkład prawdopodobieństwa na tej przestrzeni warunkowej nie zależy od wartości p . Zauważmy również, że zmienna losowa L_{2n} może przyjmować tylko parzyste wartości (ze zbioru $\{0, 2, \dots, 2n\}$). Można by przypuszczać, że jej rozkład skupia się w połowie zakresu, tj. w okolicach n . Okazuje się jednak, że zachodzi następujące, sprzeczne z intuicją twierdzenie:

Twierdzenie 4.3.

Dla każdego $k = 0, 1, \dots, n$:

$$\mathbb{P}(L_{2n} = 2k | S_{2n} = 0) = \frac{1}{n+1}$$

Dowód. Przyjmijmy w naszych rozważaniach, że $p = q = \frac{1}{2}$.

Niech $|s|, |t| < 1$. Bierzemy dwie funkcje tworzące:

$$G_{2n}(s) = \mathbb{E}(s^{L_{2n}} | S_{2n} = 0) = \sum_{k=0}^n s^{2k} \mathbb{P}(L_{2n} = 2k | S_{2n} = 0),$$

$$F_0(s) = E(s^{T_0}),$$

gdzie T_0 oznacza czas pierwszego powrotu do zera.

Teraz (znieńcka!) bierzemy następującą funkcję

$$H(s, t) = \sum_{n=0}^{\infty} t^{2n} \mathbb{P}(S_{2n} = 0) G_{2n}(s).$$

Naszym celem będzie pokazanie, że

$$G_{2n}(s) = \sum_{k=0}^n \frac{s^{2k}}{n+1}.$$

Wówczas stanie się jasne, że

$$\sum_{k=0}^n \frac{s^{2k}}{n+1} = \sum_{k=0}^n s^{2k} \mathbb{P}(L_{2n} = 2k | S_{2n} = 0),$$

a stąd (z porównania współczynników) wynika teza.

Postawimy sobie teraz cel pomocniczy (podcel), którym będzie wykazanie, iż

$$H(s, t) - 1 = \frac{1}{2} H(s, t) (F_0(t) + F_0(st)).$$

Z własności funkcji tworzącej i wartości oczekiwanej (analog wzoru na prawdopodobieństwo całkowite) otrzymujemy:

$$G_{2n}(s) = \sum_{r=1}^n \mathbb{E}(s^{L_{2n}} | S_{2n} = 0, T_0 = 2r) \mathbb{P}(T_0 = 2r | S_{2n} = 0).$$

Z jednorodności czasowej (Twierdzenie 1.3):

$$\begin{aligned} \mathbb{E}(s^{L_{2n}} | S_{2n} = 0, T_0 = 2r) &= \mathbb{E}(s^{L_{2n-2r}} | S_{2n-2r} = 0) \cdot \mathbb{E}(s^{L_{2r}} | T_0 = 2r) = \\ &= \underbrace{G_{2n-2r}(s)}_{\text{z definicji}} \cdot \sum_{k=0}^r s^{2k} \underbrace{\mathbb{P}(L_{2r} = 2k | T_0 = 2r)}_{\text{jest niezerowe tylko dla } k \in \{0, r\}} = G_{2n-2r}(s) \cdot \left(\frac{1}{2} + \frac{1}{2} s^{2r} \right). \end{aligned}$$

Z kolei z jednorodności czasowej (Twierdzenie 1.3) i wzoru na prawdopodobieństwo warunkowe:

$$\mathbb{P}(T_0 = 2r | S_{2n} = 0) = \frac{\mathbb{P}(T_0 = 2r) \mathbb{P}(S_{2n-2r} = 0)}{\mathbb{P}(S_{2n} = 0)}.$$

Po podstawieniu otrzymujemy:

$$\begin{aligned} G_{2n}(s) &= \sum_{r=1}^n G_{2n-2r}(s) \left(\frac{1}{2} + \frac{1}{2} s^{2r} \right) \frac{\mathbb{P}(T_0 = 2r) \mathbb{P}(S_{2n-2r} = 0)}{\mathbb{P}(S_{2n} = 0)} \\ \mathbb{P}(S_{2n} = 0) G_{2n}(s) &= \sum_{r=1}^n (G_{2n-2r}(s) \mathbb{P}(S_{2n-2r} = 0)) \left(\frac{1}{2} (1 + s^{2r}) \mathbb{P}(T_0 = 2r) \right) \end{aligned}$$

Po obustronnym pomnożeniu przez t^{2n} i zsumowaniu po n dostajemy:

$$\begin{aligned}
\sum_{n=1}^{\infty} t^{2n} \mathbb{P}(S_{2n} = 0) G_{2n}(s) &= \frac{1}{2} \sum_{n=1}^{\infty} \sum_{r=1}^n G_{2n-2r}(s) \mathbb{P}(S_{2n-2r} = 0) t^{2n} (1 + s^{2r}) \mathbb{P}(T_0 = 2r) \\
H(s, t) - 1 &= \frac{1}{2} \sum_{r=1}^{\infty} \sum_{n=r}^{\infty} G_{2n-2r}(s) \mathbb{P}(S_{2n-2r} = 0) t^{2n-2r} (t^{2r} + (st)^{2r}) \mathbb{P}(T_0 = 2r) = \\
&= \frac{1}{2} \sum_{r=1}^{\infty} \underbrace{\sum_{m=0}^{\infty} G_{2m}(s) \mathbb{P}(S_{2m} = 0) t^{2m}}_{=H(s,t)} (t^{2r} + (st)^{2r}) \mathbb{P}(T_0 = 2r) = \\
&= \frac{1}{2} H(s, t) \left(\sum_{r=1}^{\infty} t^{2r} \mathbb{P}(T_0 = 2r) + \sum_{r=1}^{\infty} (st)^{2r} \mathbb{P}(T_0 = 2r) \right) = \\
&= \frac{1}{2} H(s, t) (F_0(t) + F_0(st))
\end{aligned}$$

co było naszym podcelem.

Wyznaczamy:

$$\begin{aligned}
H(s, t) &= \frac{2}{2 - F_0(t) - F_0(st)} = \frac{2}{\sqrt{1-t^2} + \sqrt{1-s^2t^2}} = \\
&= \frac{2(\sqrt{1-s^2t^2} - \sqrt{1-t^2})}{t^2(1-s^2)} = \\
&= \sum_{n=0}^{\infty} t^{2n} P(S_{2n} = 0) \frac{1 - s^{2n+2}}{(n+1)(1-s^2)}.
\end{aligned}$$

Wynika stąd, że:

$$G_{2n}(s) = \frac{1}{n+1} \frac{1 - s^{2n+2}}{1 - s^2} = \frac{1}{n+1} \sum_{k=0}^n s^{2k}$$

co było naszym celem. □

4.2. Procesy gałązkowe

Siedząc w kasynie (lub w jakiejś knajpie, do której zaniósł nas stamtąd jakiś spacer losowy) i dumając nad ostatnią monetą, która została nam w kieszeni, chcielibyśmy nieraz, żeby się nam ta moneta rozmnożyła. Wiemy, że w rzeczywistości prawdopodobieństwo takiego zdarzenia jest równe zero. Lecz gdyby tak puścić wodze wyobraźni i przyjąć, że z jednej monety powstaje kilka w każdym kolejnym pokoleniu i zmienne losowe opisujące liczbę „dzieci” każdej monety są niezależne oraz o jednakowym rozkładzie, to jakie byłoby prawdopodobieństwo, że w końcu uzbieralibyśmy na ten wymarzony samochód?

Powyższy proces jest przykładem *procesu gałęzowego*, zwanym też procesem Galtona¹-Watsona². Procesy gałęzowe modelują rozwój populacji (jednopłciowej, rozmnażającej się przez podział, np. bakterii, ameb, monet, lub innych mikroorganizmów). Zmienne losowe Z_n (przyjmujące nieujemne wartości) oznaczają liczbę osobników w n -tym pokoleniu. Przyjmujemy zawsze, że jest jeden „protoplasta rodu”, czyli $Z_0 = 1$. Jak już wspomnieliśmy, zmienne losowe opisujące ile dzieci ma każdy osobnik, są niezależne o jednakowym rozkładzie (*iid*). Główne pytanie pojawiające się w tym kontekście to: jaka jest szansa, że dana populacja przeżyje?

Wprowadźmy następujące oznaczenie:

$$G_n(s) = G_{Z_n}(s) = \mathbb{E}(s^{Z_n}).$$

Zachodzi następujące twierdzenie:

Twierdzenie 4.4.

$$G_{m+n}(s) = G_n(G_m(s)) = G_m(G_n(s)) = G(G(\dots G(s)\dots)).$$

Dowód.

$$Z_{m+n} = X_1 + X_2 + \dots + X_{Z_m},$$

gdzie X_i oznacza liczbę potomków i -tego osobnika z m -tego pokolenia po n pokoleniach (czyli w $(m+n)$ -tym pokoleniu).

Na podstawie Twierdzenia 3.13:

$$G_{m+n}(s) = G_m(G_{X_1}(s)).$$

Ponadto z Twierdzenia 1.3:

$$G_{X_1}(s) = G_n(s)$$

Mamy więc:

$$G_n(s) = G_1(G_{n-1}(s)) = \dots = G_1(G_1(\dots G_1(s)\dots)),$$

a stąd już dostajemy tezę. □

¹ Francis Galton (1822-1911) — przyrodnik, antropolog i podróżnik angielski.

² Henry William Watson (1827-1903) — matematyk angielski.

5. Wykład 5: Procesy gałązkowe. Łańcuchy Markowa.

Wprowadzenie

W pierwszej części dokończymy omawianie procesów gałązkowych udowadniając lemat o wartości oczekiwanej i wariancji zmiennej losowej procesu gałązkowego. Omówimy także tzw. geometryczny proces gałązkowy i prawdopodobieństwo wyginięcia takim modelem. W drugiej części przejdziemy do zagadnienia łańcuchów Markowa przytaczając podstawowe definicje i własności.

5.1. Procesy gałązkowe (dokończenie)

Twierdzenie 4.4 dostarcza nam informacji o Z_n , ale w praktyce wyznaczenie $G_n(s)$ może okazać się trudnym zadaniem. Możemy jednak wyznaczyć momenty zmiennej losowej Z_n w funkcji wartości momentów zmiennej Z_1 .

Lemat 5.1:

Oznaczmy $\mu = \mathbb{E}(Z_1)$, $\sigma^2 = \text{Var}(Z_1)$ wtedy

$$\mathbb{E}(Z_n) = \mu^n, \text{Var}(Z_n) = \begin{cases} n\sigma^2 & \text{dla } \mu = 1, \\ \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} & \text{dla } \mu \neq 1. \end{cases}$$

Dowód. Różniczkując obustronnie równość $G_n(s) = G(G_{n-1}(s))$ w punkcie $s = 1$ otrzymujemy $\mu_n = \mathbb{E}(Z_n) = \mu \mathbb{E}(Z_{n-1})$. Rozwijając indukcyjnie równość $\mu_n = \mu \cdot \mu_{n-1}$ dostajemy $\mu_n = \mu^n$.

Dwukrotnie różniczkując tę samą równość, otrzymujemy

$$G_n''(1) = G''(1)G_{n-1}'(1)^2 + G'(1)G_{n-1}''(1)$$

korzystając z Twierdzenia 3.13 i rozwijając zależność otrzymujemy wynik. \square

5.1.1. Geometryczny proces gałązkowy

Przedstawimy pewną intuicję na temat geometrycznych procesów gałązkowych na przykładzie historycznym. Przenieśmy się do średniowiecznej Japonii. Żyli tam mężczyźni wojownicy zwani samurajami. Uważa się, że wprawny samuraj potrafił zabić przeciwnika już za pierwszym ciosem, ponoć nawet jednym cięciem przeciąć go na pół, wzdłuż od czubka głowy. Wyobraźmy sobie średniowieczną szkołę samurajów, w której każdy mistrz wybiera kolejno uczniów i przekazuje im całą swoją wiedzę. Z prawdopodobieństwem p wybrany uczeń jest

na tyle zdolny by pojąć nauki i w krótkim czasie samemu stać się mistrzem i nauczać innych. Niestety z prawdopodobieństwem $q = 1 - p$ wybrany uczeń nie jest w stanie pojąć wszystkich tajników wiedzy i przestrzegać zasad kanonu “bushido”. Zrozpaczony mistrz popełnia wtedy seppuku, a uczeń zostaje wypędzony ze szkoły. Uważa się, że samurajowie wyginęli. Pozostaje pytanie: jakie mieli szanse by przetrwać ?

Przyrost w następnej populacji w takim modelu wyznaczony jest przez funkcję:

$$\mathbb{P}(Z_1 = k) = qp^k \text{ dla } k \geq 0, \text{ gdzie } q = 1 - p$$

Wtedy funkcja tworząca prawdopodobieństwa tej zmiennej losowej jest równa $G(s) = q(1 - ps)^{-1}$, a wartość oczekiwana $\mathbb{E}(Z_1) = \frac{p}{q}$.
Indukcyjnie można pokazać, że :

$$G_n(s) = \begin{cases} \frac{n-(n-1)s}{n+1-ns} & \text{dla } p = q = \frac{1}{2}, \\ \frac{q[p^n - q^n - ps(p^{n-1} - q^{n-1})]}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} & \text{dla } p \neq q. \end{cases}$$

Obiektem naszego zainteresowania będzie prawdopodobieństwo wyginęcia społeczeństwa żyjącego w takim modelu. Oznaczmy je przez η . Wówczas:

$$\eta = \mathbb{P}(\exists n : Z_n = 0) = \mathbb{P}\left(\bigcup_{n \geq 1}^{\infty} \{Z_n = 0\}\right)$$

Ponieważ ciąg zdarzeń mówiących o wyginieciu w danej chwili czasu tworzy ciąg wstępujący ($\{Z_n = 0\} \Rightarrow \{Z_{n+1} = 0\}$), to ze skończonej addytywności otrzymujemy

$$\eta = \mathbb{P}\left(\bigcup_{n \geq 1}^{\infty} \{Z_n = 0\}\right) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0)$$

Jakie jest prawdopodobieństwo, że konkretne $Z_n = 0$?

$$\mathbb{P}(Z_n = 0) = G_n(0) = \begin{cases} \frac{n}{n+1} & \text{dla } p = q, \\ \frac{q(p^n - q^n)}{p^{n+1} - q^{n+1}} & \text{dla } p \neq q. \end{cases}$$

Przy przejściu z $n \rightarrow \infty$ mamy:

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0) = \begin{cases} 1 & \text{dla } p \leq q, \\ \frac{q}{p} & \text{dla } p > q. \end{cases}$$

Twierdzenie 5.2 (O wyginieciu).

- (a) Prawdopodobieństwo wyginiecia η jest równe najmniejszemu nieujemnemu pierwiastkowi równania $s = G(s)$.
(b) W szczególności

$$\eta = \begin{cases} 1 & \text{gdy } \mu < 1 \text{ lub } \mu = 1 \text{ i } \sigma > 0, \\ < 1 & \text{gdy } \mu > 1 \text{ lub } \mu = 1 \text{ i } \sigma = 0. \end{cases}$$

Uwaga 5.3. Zauważmy, że $\sigma = 0$ implikuje $\mathbb{P}(Z_1 = \mu) = 1$, co daje $\eta = 0$ bo nikt nie może mieć zerowej liczby potomków.

Dowód twierdzenia 5.2.

(a) Niech $\eta_n = \mathbb{P}(Z_n = 0)$ wtedy

$$\forall_n \eta_n = \mathbb{P}(Z_n = 0) = G_n(0) \stackrel{4.4}{=} G(G_{n-1}(0)) = G(\eta_{n-1})$$

η jest pierwiastkiem ponieważ

$$\eta = \lim_{n \rightarrow \infty} \eta_n = \lim_{n \rightarrow \infty} G(\eta_{n-1}) \stackrel{\text{ciągłość } G}{=} G(\lim_{n \rightarrow \infty} \eta_{n-1}) = G(\eta)$$

Pokażemy, że η jest najmniejszym pierwiastkiem. Niech ψ będzie nieujemnym pierwiastkiem równania $s = G(s)$. Wtedy, ponieważ G jest niemalejąca

$$\eta_1 = G(0) \leq G(\psi) = \psi$$

indukcyjnie

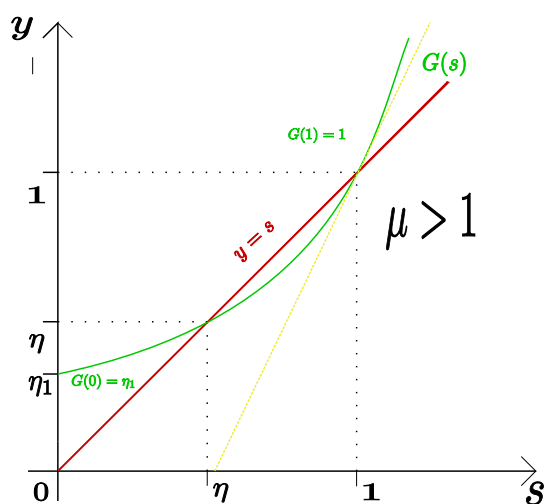
$$\eta_2 = G(\eta_1) \leq G(\psi) = \psi$$

$$\forall_{1 < i \leq n} \eta_i = G(\eta_{i-1}) \leq G(\psi) = \psi$$

stąd

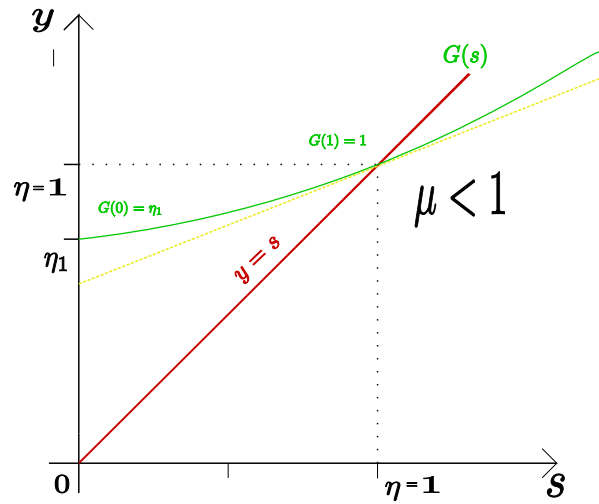
$$\forall_n \eta_n \leq \psi \Rightarrow \lim_{n \rightarrow \infty} \eta_n \leq \psi$$

(b) Przypadek (I) $\mu = G'(1) > 1$



Rysunek 5.1. przypadek $\mu < 1$

Na Rysunku 5.1.1 μ jest współczynnikiem stycznej do $G(s)$ w punkcie $s = 1$. Jak widać, w tym przypadku funkcja $G(s)$ musi przecinać się z wykresem funkcji $y = s$ poniżej punktu $y = 1$ (ponieważ $G(0) = \eta_1$) stąd z (a) $\eta < 1$. W przypadku gdy $\eta_1 = 0$ mamy $\eta = 0$.
Przypadek (II) $\mu = G'(1) < 1$



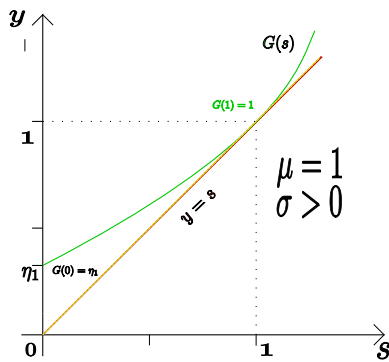
Rysunek 5.2. przypadek $\mu < 1$

Na Rysunku 5.1.1 μ jest współczynnikiem stycznej do $G(s)$ w punkcie $s = 1$. Jak widać, w tym przypadku $s = 1$ jest najmniejszym pierwiastkiem równania $G(s) = s$. Dla $\mu < 1$ z nierówności Markowa mamy:

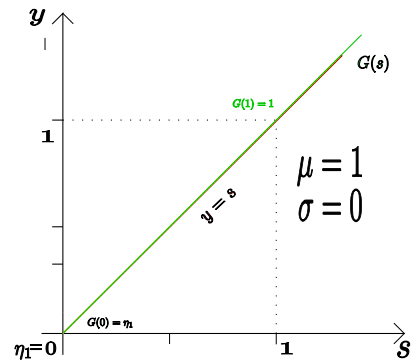
$$\mathbb{P}(Z_n > 0) \leq \mathbb{E}(Z_n) = \mu^n \text{ ale } \mu^n \rightarrow 0 \text{ bo } \mu < 1$$

Czyli $\mathbb{P}(Z_n = 0) \rightarrow 1$, a stąd $\eta = 1$.

Przypadek (III) $\mu = 1$



Rysunek 5.3. $\mu = 1$, $\sigma > 0$



Rysunek 5.4. $\mu = 1$, $\sigma = 0$

□

Przykład 5.4 (Zastosowania twierdzenia o wyginieciu dla geometrycznego procesu gałazkowego.). Szukamy rozwiązań równania $G(s) = s$.

$$G(s) = q(1 - ps)^{-1} = s$$

$$q = s - ps^2$$

$$ps^2 - s + q = 0$$

$$\Delta = 1 - 4pq$$

$$s_{1,2} = \frac{1 \pm |p - q|}{2p}$$

Stąd

$$s_1 = 1 \text{ dla } p \leq q, \quad s_2 = \frac{q}{p} \text{ dla } p > q$$

Jak widać uzyskaliśmy taki sam wynik jak przy bezpośrednim policzeniu prawdopodobieństwa wyginięcia dla tego procesu. Ten sam rezultat uzyskujemy także przy przeliczaniu prawdopodobieństwa dojścia do 0 w spacerze losowym startującym z 1 o prawdopodobieństwie q ruchu w lewo i prawdopodobieństwie p pójścia w prawo.

Możemy wyobrazić sobie następującą analogię: startujemy z 1 (w naszym społeczeństwie mamy do czynienia z pojedynczą jednostką która mieszka w jednym domostwie). Każdy nowy potomek potrzebuje swojego własnego mieszkania, budujemy mu więc nowy dom. Nasz spacer losowy będzie liczył liczbę domów w wiosce. Rozpatrujemy każdego „rodzica” po kolei względem wieku. Kiedy wyróżniony osobnik umiera, po jego śmierci domostwo zostaje wyburzone – z każdym pogrzebem więc liczba domostw w naszej wiosce zmniejsza się o jeden. Wyginięcie populacji oznaczać będzie, że w naszej wiosce nie pozostał żaden dom, co jest tożsame z dojściem do 0 w spacerze losowym.

5.2. Łańcuchy Markowa

Procesem Markowa nazywamy ciąg zmiennych losowych, w którym prawdopodobieństwo tego co się zdarzy zależy jedynie od stanu obecnego. W tym rozdziale zajmiemy się zagadnieniem łańcuchów Markowa, czyli procesów Markowa zdefiniowanych na dyskretnej przestrzeni stanów. Łatwo zauważyć, iż rozpatrywane poprzednio zagadnienia spacerów losowych, czy też procesów gałązkowych, są przykładami procesów Markowa.

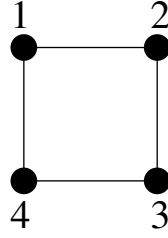
Oznaczmy przez $X = (X_0, X_1, \dots)$ ciąg zmiennych losowych dyskretnych. Wartość zmiennej X_n będziemy nazywać stanem łańcucha w chwili n . Zakładamy, że zbiór stanów S jest przeliczalny.

Definicja 5.5. Mówimy, że X jest *łańcuchem Markowa* gdy spełnia *własność Markowa*. Tzn.

$$\forall_{n \geq 1} \forall_{s, X_0, X_1, \dots, X_{n-1} \in S} \mathbb{P}(X_n = s | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = s | X_{n-1} = x_{n-1})$$

Przykład 5.6. Wyobraźmy sobie, że jesteśmy na Florydzie i spacerujemy po bagnach, w których czają się bardzo głodne aligatory. W trosce o własne życie musimy więc przemieszczać się po ściśle wyznaczonych (przez drewniane bale) ścieżkach wytyczonych pomiędzy kolejnymi punktami. Na każdym rozwidleniu w sposób losowy podejmujemy decyzję, w którym kierunku się udamy.

Spójrzmy na przykładowe rozmieszczenie takich bal i ścieżek na Rysunku 5.6. Oznaczmy przez X_n indeks wierzchołka, w którym znajdziemy się w czasie n . Wtedy (X_0, X_1, \dots)



Rysunek 5.5. Przykładowy graf ścieżek i połączeń

będzie ciągiem zmiennych losowych o wartościach ze zbioru $\{1, 2, 3, 4\}$. Załóżmy, że startujemy w punkcie 1. Mamy wówczas $\mathbb{P}(X_0 = 1) = 1$. Następnie z prawdopodobieństwem $\frac{1}{2}$ przemieszczamy się do kolejnych wierzchołków:

$$\mathbb{P}(X_1 = 2) = \frac{1}{2} \text{ oraz } \mathbb{P}(X_1 = 4) = \frac{1}{2}$$

Załóżmy, że w czasie n znajdziemy się w wierzchołku 2. Otrzymamy wtedy:

$$\mathbb{P}(X_{n+1} = 1|X_n = 2) = \frac{1}{2} \text{ oraz } \mathbb{P}(X_{n+1} = 3|X_n = 2) = \frac{1}{2}$$

Zauważmy teraz, że gdybyśmy rozpatrzyli całą historię spaceru, otrzymane prawdopodobieństwa nie zmieniają się:

$$\mathbb{P}(X_{n+1} = 1|X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = 2) = \frac{1}{2}$$

oraz

$$\mathbb{P}(X_{n+1} = 3|X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = 2) = \frac{1}{2}$$

bez względu na wybór i_0, \dots, i_{n-1} . Jest to znak że podany proces spełnia warunek Markowa.

Dla każdego łańcucha Markowa można określić tzw. macierz przejść.

Definicja 5.7. Niech P będzie macierzą o wymiarach $(k \times k)$ i elementach $\{p_{ij} : i, j = 1, \dots, k\}$. Ciąg zmiennych losowych (X_0, X_1, \dots) o wartościach ze skończonego zbioru stanów $S = \{s_1, \dots, s_k\}$ nazywamy procesem Markowa z *macierzą przejść* P , jeżeli dla każdego n , dowolnych $i, j \in \{1, \dots, k\}$ i wszystkich $i_0, \dots, i_{n-1} \in \{1, \dots, k\}$ mamy:

$$\mathbb{P}(X_{n+1} = s_j|X_0 = s_{i_0}, X_1 = s_{i_1}, \dots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) = \mathbb{P}(X_{n+1} = s_j|X_n = s_i) = p_{ij}$$

W przykładzie 5.6 macierz przejścia przyjmie postać:

$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

Obserwacja 5.8. Elementy macierzy przejść $(p_{ij})_{i,j=1,\dots,k}$ spełniają następujące własności:

- (i) $\forall_n p_{ij} = \mathbb{P}(X_{n+1} = s_j | X_n = s_i)$
- (ii) $p_{ij} \geq 0$ dla każdych $i, j \in \{1, \dots, k\}$
- (iii) $\forall_i \sum_j p_{ij} = 1$

Zauważmy, że na postać łańcucha Markowa duży wpływ ma miejsce rozpoczęcia procesu i rozkład prawdopodobieństwa dla punktu wyjścia.

Definicja 5.9. Łańcuch Markowa nazywamy *jednorodnym* gdy w każdej chwili czasu opisuje go ta sama macierz przejść. Łańcuch nazywamy *niejednorodnym* w przypadku gdy opisuje go ciąg niejednakowych macierzy.

Definicja 5.10. Macierz przejść w n krokach z m do $m + n$ definiujemy następująco:

$$P(m, m + n) = (p_{ij}(m, m + n))_{(k \times k)}$$

gdzie

$$p_{ij}(m, m + n) = \mathbb{P}(X_{m+n} = j | X_m = i)$$

Obserwacja 5.11. Z przypadku gdy łańcuch jest jednorodny możemy przyjąć, że $P(m, m + 1) = P$.

Twierdzenie 5.12 (Chapmana - Kołmogorowa).

Dla jednorodnych łańcuchów Markowa zachodzi równość:

$$\forall_n \forall_m P(m, m + n) = P^n$$

Dowód. Korzystamy ze wzoru na prawdopodobieństwo całkowite w przestrzeni warunkowej

$$p_{ij}(m, m + n + r) \stackrel{5.10}{=} \underbrace{\sum_{k \in S} p_{ik}(m, m + n) p_{kj}(m + n, m + n + r)}_{\text{tworzą układ zupełny zdarzeń}}$$

Powyzsza równość to nic innego jak mnożenie macierzy, więc:

$$P(m, m + n + r) = P(m, m + n) \cdot P(m + n, m + n + r) = P^n \cdot P^r = P^{n+r}$$

Cofając się indukcyjnie otrzymujemy:

$$P(m, m + n) = P^n$$

□

6. Wykład 6: Klasyfikacja łańcuchów Markowa. Rozkłady stacjonarne.

Wstęp

Na ostatnim wykładzie dowiedzieliśmy się czym są łańcuchy Markowa oraz poznaliśmy podstawowe zagadnienia z nimi związane (jednorodność, macierze przejść, Twierdzenie Chapmana-Kołmogorowa). W pierwszej części tego wykładu omówimy zagadnienie stanu początkowego oraz dokonamy fragmentarycznej klasyfikacji łańcuchów Markowa ze względu na relacje zachodzące między poszczególnymi stanami. Na końcu zajmiemy się badaniem rozkładu stacjonarnego oraz przytoczymy twierdzenie o zbieżności łańcuchów.

6.1. Rozkład początkowy

W badaniu łańcuchów Markowa szczególną rolę odgrywa stan początkowy – zauważmy, że może on wpłynąć na to, iż do niektórych stanów nie dotrzemy nigdy, a do innych jedynie po pewnym czasie. Formalnie rzecz biorąc stan początkowy to po prostu zmienna losowa X_0 . Nie należy się więc dziwić, że czasem łańcuch Markowa nie będzie zaczynał się z jednego wyznaczonego stanu, ale z pewnego rozkładu prawdopodobieństwa na przestrzeni stanów.

Definicja 6.1. *Rozkładem początkowym* nazywać będziemy wektor μ_0 określony w następujący sposób:

$$\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_k^{(0)}) = (\mathbb{P}(X_0 = s_1), \mathbb{P}(X_0 = s_2), \dots, \mathbb{P}(X_0 = s_k)),$$

gdzie $\{s_1, \dots, s_k\}$ jest zbiorem stanów łańcucha.

Ponieważ $\mu^{(0)}$ jest rozkładem prawdopodobieństwa, więc

$$\sum_{i=1}^k \mu_i^{(0)} = 1.$$

Analogicznie przez $\mu^{(1)}, \mu^{(2)}, \dots$ możemy oznaczyć rozkłady prawdopodobieństwa w momentach czasu 1, 2, ... (czyli rozkłady zmiennych X_1, X_2, \dots) tak, że

$$\mu^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \dots, \mu_k^{(n)}) = (\mathbb{P}(X_n = s_1), \mathbb{P}(X_n = s_2), \dots, \mathbb{P}(X_n = s_k))$$

Przykład 6.2. Wróćmy na chwilę do Przykładu 5.6. Był tam dany spacer losowy po kwadracie (Rysunek 5.6), który zawsze zaczynaliśmy od wierzchołka 1. Jego rozkład początkowy ma zatem postać $\mu^{(0)} = (1, 0, 0, 0)$. Ponadto można łatwo zauważyć, że w drugim kroku rozkład tego łańcucha ma postać:

$$\mu^{(1)} = \left(\frac{1}{2}, 0, \frac{1}{2}, 0\right).$$

Okazuje się, że jeśli mamy jednorodny łańcuch Markowa (zobacz Definicję 5.9), znamy jego rozkład początkowy $\mu^{(0)}$ oraz macierz przejść P , to w prosty sposób możemy wyznaczyć wszystkie rozkłady $\mu^{(1)}, \mu^{(2)}, \dots$. Mówi o tym następujące twierdzenie:

Twierdzenie 6.3.

Jeśli (X_1, X_2, \dots) jest jednorodnym łańcuchem Markowa o skończonym zbiorze stanów $\{s_1, \dots, s_k\}$, rozkładzie początkowym $\mu^{(0)}$ i macierzy przejść P , to

$$\forall_n \mu^{(n)} = \mu^{(0)} P^n$$

Dowód. Rozważmy przypadek kiedy $n = 1$. Dla $j = 1, \dots, k$, otrzymujemy

$$\begin{aligned} \mu_j^{(1)} &= \mathbb{P}(X_1 = s_j) = \sum_{i=1}^k \mathbb{P}(X_0 = s_i, X_1 = s_j) \\ &= \sum_{i=1}^k \mathbb{P}(X_0 = s_i) \mathbb{P}(X_1 = s_j \mid X_0 = s_i) \\ &= \sum_{i=1}^k \mu_i^{(0)} p_{i,j} = (\mu^{(0)} P)_j \end{aligned}$$

gdzie $(\mu^{(0)} P)_j$ oznacza j -ty element wektora $\mu^{(0)} P$. Stąd $\mu^{(1)} = \mu^{(0)} P$.

Przypadek ogólny udowodnimy indukcyjnie. Ustalmy m i założmy, że teza twierdzenia jest prawdziwa dla $n = m$. Dla $n = m + 1$ mamy zatem

$$\begin{aligned} \mu_j^{(m+1)} &= \mathbb{P}(X_{m+1} = s_j) = \sum_{i=1}^k \mathbb{P}(X_m = s_i, X_{m+1} = s_j) \\ &= \sum_{i=1}^k \mathbb{P}(X_m = s_i) \mathbb{P}(X_{m+1} = s_j \mid X_m = s_i) \\ &= \sum_{i=1}^k \mu_i^{(m)} p_{i,j} = (\mu^{(m)} P)_j, \end{aligned}$$

a stąd $\mu^{(m+1)} = \mu^{(m)} P$. Z założenia indukcyjnego wiemy, że $\mu^{(m)} = \mu^{(0)} P^m$, skąd ostatecznie otrzymujemy

$$\mu^{(m+1)} = \mu^{(m)} P = \mu^{(0)} P^m P = \mu^{(0)} P^{(m+1)}$$

□

Powyższe twierdzenie łatwo daje się uogólnić na przypadek łańcucha niejednorodnego:

Twierdzenie 6.4.

Niech (X_1, X_2, \dots) będzie niejednorodnym łańcuchem Markowa o skończonym zbiorze stanów, rozkładzie początkowym $\mu^{(0)}$ i macierzach przejść $P^{(1)}, P^{(2)}, \dots$.

Wówczas

$$\forall_n \mu^{(n)} = \mu^{(0)} P^{(1)} P^{(2)} \dots P^{(n)}.$$

Dowód. Dowód jest analogiczny jak w przypadku Twierdzenia 6.3.

□

6.2. Przykłady

W tej części wykładu pokażemy kilka przykładów obrazujących pewne procesy Markowa przedstawione za pomocą macierzy przejść.

Przykład 6.5 (Pogoda w Göteborgu). Wyobraźmy sobie, że jesteśmy w szwedzkim mieście Göteborg i chcemy przewidzieć jaka pogoda będzie jutro. Niektórzy mówią, że najlepszym sposobem jest po prostu przypuścić, że będzie ona taka sama jak dziś. Gdyby przyjąć, że stwierdzenie to jest prawdziwe, to naturalne wydaje się modelowanie pogody za pomocą łańcuchów Markowa. Ponieważ ma to być prosty przykład, to założymy, że są możliwe tylko dwa rodzaje pogody: słońce lub deszcz. Jeśli powyższy sposób przewidywania pogody sprawdza się w 75% czasu (niezależnie czy dziś świeci słońce czy pada deszcz), to pogoda tworzy łańcuch Markowa ze zbiorem stanów $S = \{s_1, s_2\}$, gdzie $s_1 = \text{„deszcz”}$, $s_2 = \text{„słońce”}$ i macierzą przejść

$$P = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}.$$

Przykład 6.6 (Pogoda w Los Angeles). W poprzednim przykładzie dla pogody w Göteborgu zachodziła symetria między „deszczem” i „słońcem”. Jednak w Los Angeles słońce świeci znacznie częściej niż pada deszcz, dlatego bardziej uzasadniony będzie model łańcucha Markowa z macierzą przejść

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{10} & \frac{9}{10} \end{bmatrix}$$

Przykład 6.7 (Internet). Okazuje się, że przy pewnych dodatkowych założeniach, również na poruszanie się po internecie można patrzeć jak na proces Markowa. Jeśli przez S oznaczymy zbiór wszystkich stron internetowych w danym momencie (jest on oczywiście skończony), a przez d_i liczbę odnośników do innych witryn umieszczonych na i -tej stronie, to otrzymamy proces Markowa o macierzy przejść P takiej, że

$$p_{i,j} = \begin{cases} \frac{1}{d_i} & \text{gdy na stronie } s_i \text{ istnieje link do strony } s_j \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Gdybyśmy w przykładzie tym dopuścili również używania przycisku powrotu do poprzedniej strony, to proces straciłby własność Markowa, ponieważ to, gdzie znajdziemy się w następnym kroku, zależałoby również od tego, które strony odwiedziliśmy wcześniej. Zatem przyszłość nie zależałaby tylko od terażniejszości, lecz także od przeszłości.

Przykład 6.8 (Dokładniejsza pogoda w Göteborgu). W Przykładzie 6.5 rozważaliśmy mało rzeczywisty przypadek, ponieważ pogoda była niezależna od pory roku. Teraz rozbudujemy trochę tę sytuację. Do zbioru stanów dodamy stan s_3 oznaczający „śnieg” oraz wprowadzimy dwie różne macierze – będzie to przykład niejednorodnego łańcucha Markowa. Zmiany pogody w lecie (np. od maja do września) ilustrować będzie macierz P_{lato} , a w zimie (tj. od października do kwietnia) macierz P_{zima} :

$$P_{lato} = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \quad P_{zima} = \begin{bmatrix} \frac{1}{2} & \frac{3}{10} & \frac{1}{5} \\ \frac{3}{20} & \frac{7}{10} & \frac{3}{20} \\ \frac{1}{5} & \frac{3}{10} & \frac{1}{2} \end{bmatrix}$$

6.3. Klasyfikacja łańcuchów Markowa

W rozdziale tym omówimy dwie ważne własności łańcuchów Markowa: okresowość i rozkładalność. Są one kluczowe dla badania rozkładu stacjonarnego, któremu poświęcony będzie kolejny rozdział. Poniższe definicje wprowadzamy dla łańcucha Markowa (X_0, X_1, \dots) o wartościach w skończonym zbiorze stanów $\{s_1, \dots, s_k\}$ i macierzy przejść P .

Definicja 6.9. Mówimy, że stan s_i *komunikuje się* ze stanem s_j , co zapisujemy $s_i \rightarrow s_j$, jeśli istnieje n takie, że

$$\mathbb{P}(X_{m+n} = s_j \mid X_m = s_i) > 0.$$

Intuicyjnie oznacza to, że ze stanu s_i da się dojść do stanu s_j z pewnym dodatnim prawdopodobieństwem.

Obserwacja 6.10. Dla łańcuchów jednorodnych wspomniane prawdopodobieństwo komunikowania się jest niezależne od m i równe $(P^n)_{i,j}$ (element i, j -ty n -tej potęgi macierzy przejść).

Definicja 6.11. Jeśli $s_i \rightarrow s_j$ i $s_j \rightarrow s_i$, to mówimy, że stany *komunikują się ze sobą*, co zapisujemy $s_i \leftrightarrow s_j$.

Definicja 6.12. Łańcuch Markowa nazywamy nieredukowalnym, jeśli

$$\forall_{s_i, s_j \in S} (s_i \leftrightarrow s_j)$$

Łańcuch, który nie jest nieredukowalny nazywamy będziemy redukowalnym.

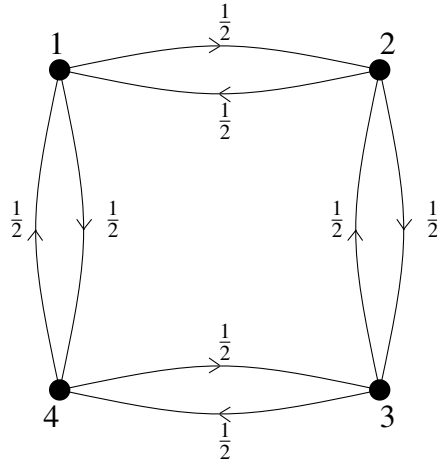
Intuicyjnie nieredukowalność łańcucha oznacza, że z każdego stanu można osiągnąć po jakimś czasie dowolny inny stan z pewnym dodatnim prawdopodobieństwem.

Wygodnym sposobem przedstawiania łańcuchów Markowa są tzw. grafy przejść.

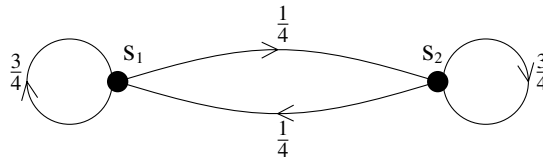
Definicja 6.13. *Grafem przejść* łańcucha Markowa nazywamy skierowany graf ważony $G = (V, E)$, w którym zbiorem wierzchołków jest zbiór stanów łańcucha, $(V = S)$, a dwa wierzchołki v i w są połączone krawędzią o wadze p , gdy $p = \mathbb{P}(X_1 = w \mid X_0 = v) > 0$ (wierzchołek w jest osiągalny z wierzchołka v w jednym kroku).

Obserwacja 6.14. Jeśli chcemy w prosty sposób sprawdzić czy łańcuch Markowa jest nieredukowalny, wystarczy spojrzeć na jego graf przejść i sprawdzić czy istnieje ścieżka skierowana między dowolnymi dwoma wierzchołkami.

Przykład 6.15. Graf przejść związany ze spacerem losowym po kwadracie z Przykładu 5.6 będzie wyglądał tak, jak na Rysunku 6.3.



Rysunek 6.1. Graf przejść spaceru losowego z z Przykładu 5.6.



Rysunek 6.2. Graf przejść pogody w Göteborgu z Przykładu 6.5.

Przykład 6.16. Graf przejść związany z pogodą w Göteborgu z Przykładu 6.5 będzie wyglądał tak, jak na Rysunku 6.3.

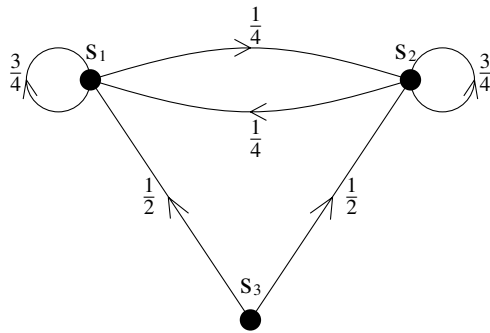
Przykład 6.17. Grafy przejść związany z dokładniejszą pogodą w Göteborgu z Przykładu 6.8 będą wyglądały tak, jak na Rysunkach 6.3 (lato) i 6.3 (zima). Brak łuku (krawędzi w grafie) oznacza, że prawdopodobieństwo przejścia między dwoma stanami jest zerowe.

Dla każdego skończonego lub nieskończonego zbioru $\{a_1, a_2, \dots\}$ liczb naturalnych przez $NWD\{a_1, a_2, \dots\}$ oznaczamy największy wspólny dzielnik liczb a_1, a_2, \dots

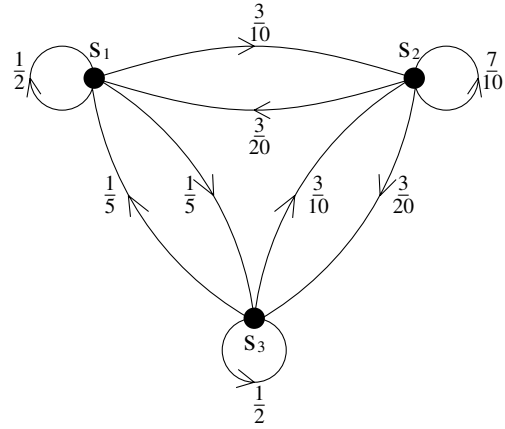
Definicja 6.18. Okres $d(s_i)$ stanu $s_i \in S$ definiujemy następująco:

$$d(s_i) = NWD\{n \geq 1 : (P^n)_{i,i} > 0\}.$$

Intuicyjnie oznacza to NWD ze wszystkich czasów, w których możemy wrócić do tego samego miejsca.



Rysunek 6.3. Graf przejść dokładniejszej pogody w Göteborgu latem z Przykładu 6.5.



Rysunek 6.4. Graf przejść dokładniejszej pogody w Göteborgu zimą z Przykładu 6.5.

Definicja 6.19. Jeśli $d(s_i) = 1$, to mówimy, że stan s_i jest *nieokresowy*. Łańcuch Markowa nazywamy *nieokresowym*, jeśli wszystkie jego stany są nieokresowe. Łańcuch, który nie jest nieokresowy nazywamy *okresowym*.

Obserwacja 6.20. Łańcuch Markowa, w którym dla każdego stanu istnieje pewne niezerowe prawdopodobieństwo pozostania w nim, jest w sposób naturalny łańcuchem nieokresowym.

Przykład 6.21. Łańcuch z Przykładu 5.6 jest okresowy, gdyż okres każdego stanu wynosi 2. Łańcuchy z przykładów 6.5, 6.6, 6.8 są nieokresowe.

Poniżej przedstawiamy pewien lemat z teorii liczb, który wykorzystamy w dowodzie kolejnego twierdzenia.

Lemat 6.22:

Jeśli $A = \{a_1, a_2, \dots\}$ jest podzbiorem zbioru liczb naturalnych spełniającym warunki:

- (i) $NWD\{a_1, a_2, \dots\} = 1$
 - (ii) A jest zamknięty ze względu na dodawanie (tzn. jeśli $a \in A$ i $a' \in A$, to $a + a' \in A$),
- to istnieje $N < \infty$ takie, że $n \in A$ dla każdego $n \geq N$.

Dowód. Oznaczmy dla wygody $b_i = NWD(a_1, \dots, a_i)$. Oczywiście jest, że dla każdego $i \in \mathbb{N}$ mamy $b_i \geq b_{i+1}$ oraz $b_i \geq 1$. Otrzymujemy w ten sposób nierosnący ciąg liczb naturalnych $(b_i)_{i \in \mathbb{N}}$, dla którego $\exists_k \forall_{i \geq k} b_i = 1$. Istotnie, gdyby $\forall_i b_i > 1$, to z zasady minimum otrzymalibyśmy, że $\exists_{a > 1} \exists_{I \in \mathbb{N}} \forall_{i > I} b_i = a$, ale wówczas mielibyśmy $NWD\{a_1, a_2, \dots\} = a > 1$, co przeczy założeniu. Istnieje więc $\{n_1, \dots, n_k\} \subset A$ taki, że $NWD(n_1, \dots, n_k) = 1$. Z podstawowego faktu teorii liczb wiadomo, że

$$\exists_{l_1, \dots, l_k \in \mathbf{Z}} \sum_{i=1}^k l_i n_i = 1.$$

Niech $N = n_1 \sum_{i=1}^k |l_i| n_i$. Wtedy $n \geq N$ ma przedstawienie postaci $n = N + ln_1 + m$, gdzie $l \geq 0$, $m < n_1$. Stąd

$$n = n_1 \sum_{i=1}^k |l_i| n_i + ln_1 + m \sum_{i=1}^k l_i n_i = ln_1 + \sum_{i=1}^k (n_1 |l_i| + ml_i) n_i,$$

czyli $n = c_1 n_1 + \dots + c_k n_k$, gdzie $c_i \geq 0$. Ponieważ $n_1, \dots, n_k \in A$, to z addytywności zbioru A mamy, że $\forall_i c_i n_i \in A$, więc również $n \in A$. Udowodniliśmy zatem, że $\exists_{N \in \mathbb{N}} \forall_{n \geq N} n \in A$, co kończy dowód lematu. \square

Następujące twierdzenie mówi nam o tym, że jeśli łańcuch Markowa jest nieokresowy, to istnieje taki moment czasu, od którego zawsze (z punktu widzenia chwili obecnej) będziemy mogli wrócić do tego samego miejsca z pewnym dodatnim prawdopodobieństwem.

Twierdzenie 6.23.

Jeśli (X_1, X_2, \dots) jest nieokresowym łańcuchem Markowa o skończonym zbiorze stanów $S = \{s_1, \dots, s_k\}$ i macierzy przejść P , to istnieje $N < \infty$ takie, że dla każdego $i \in \{1, \dots, k\}$ i dla każdego $n \geq N$ zachodzi

$$(P^n)_{i,i} > 0.$$

Dowód. Niech $A_i = \{n \geq 1 : (P^n)_{ii} > 0\}$ dla każdego $i \in \{1, \dots, k\}$ (tzn. niech A_i będzie zbiorem możliwych czasów powrotu do s_i , startując z s_i). Z założenia, że łańcuch Markowa jest nieokresowy wiemy, że każdy stan s_i jest nieokresowy, a stąd wynika, że $NWD(A_i) = 1$ dla każdego $i \in \{1, \dots, k\}$. Co więcej, jeśli $a, a' \in A_i$, to $\mathbb{P}(X_a = s_i | X_0 = s_i) > 0$ i $\mathbb{P}(X_{a+a'} = s_i | X_a = s_i) > 0$. Stąd

$$\begin{aligned} \mathbb{P}(X_{a+a'} = s_i | X_0 = s_i) &\geq \mathbb{P}(X_a = s_i, X_{a+a'} = s_i | X_0 = s_i) = \\ &= \mathbb{P}(X_a = s_i | X_0 = s_i) \mathbb{P}(X_{a+a'} = s_i | X_a = s_i) > \\ &> 0, \end{aligned}$$

co oznacza, że $a + a' \in A_i$, czyli A_i jest zamknięty ze względu na dodawanie.

A_i spełnia więc założenia Lematu 6.22, z którego wynika, że istnieje $N_i < \infty$ takie, że $(P^n)_{i,i} > 0$ dla każdego $n \geq N_i$. Możemy więc wziąć $N = \max_{i=1, \dots, k} N_i$, które spełnia tezę naszego twierdzenia. \square

Poniższy wniosek zostanie wykorzystany w dowodzie Twierdzenia Ergodecznego o zbieżności łańcuchów Markowa (Twierdzenie 6.30).

Wniosek 6.24. Jeśli (X_1, X_2, \dots) jest nieredukowalnym i nieokresowym łańcuchem Markowa o skończonym zbiorze stanów $S = \{s_1, \dots, s_k\}$ i macierzy przejścia P , to istnieje $M < \infty$ takie, że $(P^n)_{i,j} > 0$ dla każdego $i, j \in \{1, \dots, k\}$ i każdego $n \geq M$.

Dowód. Z założenia, że łańcuch jest nieokresowy i z Twierdzenia 6.23 istnieje $N < \infty$ takie, że $(P^n)_{i,i} > 0$, dla każdego $i \in \{1, \dots, k\}$ i dla każdego $n \geq N$. Niech $s_i, s_j \in S$ będą

dowolnymi, ale ustalonymi stanami. Z założenia, że łańcuch jest nieredukowalny możemy znaleźć $n_{i,j}$ takie, że $(P^{n_{i,j}})_{i,j} > 0$. Niech $M_{i,j} = N + n_{i,j}$. Dla każdego $m \geq M_{i,j}$, mamy

$$\begin{aligned} \mathbb{P}(X_m = s_j \mid X_0 = s_i) &= \mathbb{P}(X_{m-n_{i,j}} = s_i, X_m = s_j \mid X_0 = s_i) = \\ &= \mathbb{P}(X_{m-n_{i,j}} = s_i \mid X_0 = s_i) \mathbb{P}(X_m = s_j \mid X_{m-n_{i,j}} = s_i) > \\ &> 0 \end{aligned}$$

(indeks $m - n_{i,j}$ jest dodatni ponieważ $m - n_{i,j} \geq M - n_{i,j} = N$). Pokazaliśmy zatem, że $(P^m)_{i,j} > 0$ dla każdego $m \geq M_{i,j}$. Możemy więc wziąć $M = \max_{i,j=1,\dots,k} M_{i,j}$, które spełnia tezę naszego twierdzenia. \square

6.4. Rozkład stacjonarny

Definicja 6.25. Niech (X_1, X_2, \dots) będzie łańcuchem Markowa o skończonym zbiorze stanów $S = \{s_1, \dots, s_k\}$ i macierzy przejścia P . Wektor $\pi = (\pi_1, \dots, \pi_k)$ nazywamy *rozkładem stacjonarnym* łańcucha Markowa, jeśli spełnia warunki:

- (i) $\pi_i \geq 0$ dla $i = 1, \dots, k$ oraz $\sum_{i=1}^k \pi_i = 1$,
- (ii) $\pi P = \pi$, tzn. $\sum_{i=1}^k \pi_i p_{i,j} = \pi_j$ dla $j = 1, \dots, k$.

W powyższej definicji warunek pierwszy mówi nam, że π musi być rozkładem prawdopodobieństwa na $\{s_1, \dots, s_k\}$. Z warunku drugiego wynika, że jeśli rozkład początkowy $\mu^{(0)} = \pi$, to

$$\mu^{(1)} = \mu^{(1)} P = \pi P = \pi,$$

a przez iterację otrzymujemy $\mu^{(n)} = \pi$ dla każdego n .

W dalszej części tego rozdziału rozważać będziemy zagadnienia istnienia i unikalności rozkładu stacjonarnego oraz zbadamy zbieżność do rozkładu stacjonarnego przy dowolnym rozkładzie początkowym. Musimy więc określić co znaczy, że ciąg rozkładów prawdopodobieństwa $\nu^{(1)}, \nu^{(2)}, \dots$ zbiega do rozkładu ν . W tym celu wprowadzimy przestrzeń metryczną określoną na rozkładach prawdopodobieństwa z metryką zwaną *total variation distance*.

Definicja 6.26. Jeśli $\nu^{(1)} = (\nu_1^{(1)}, \dots, \nu_k^{(1)})$ i $\nu^{(2)} = (\nu_1^{(2)}, \dots, \nu_k^{(2)})$ są rozkładami prawdopodobieństwa na $S = \{s_1, \dots, s_k\}$, to *total variation distance* między $\nu^{(1)}$ i $\nu^{(2)}$ definiujemy jako

$$d_{TV}(\nu^{(1)}, \nu^{(2)}) = \frac{1}{2} \sum_{i=1}^k |\nu_i^{(1)} - \nu_i^{(2)}|.$$

Jeśli $\nu^{(1)}, \nu^{(2)}, \dots$ są rozkładami prawdopodobieństwa na S , to mówimy, że $\nu^{(n)}$ zbiega do ν w sensie całkowitej wariancji gdy $n \rightarrow \infty$, jeśli

$$\lim_{n \rightarrow \infty} d_{TV}(\nu^{(n)}, \nu) = 0$$

Piszemy wtedy $\nu^{(n)} \xrightarrow{TV} \nu$.

Odległość w tej przestrzeni metrycznej posiada naturalną interpretację, którą wyraża równość:

$$d_{TV}(\nu^{(1)}, \nu^{(2)}) = \max_{A \subseteq S} | \nu^{(1)}(A) - \nu^{(2)}(A) | .$$

Dla dalszych rozważań niezbędne jest wprowadzenie nowej definicji:

Definicja 6.27. Jeśli łańcuch Markowa (X_0, X_1, \dots) o skończonym zbiorze stanów $S = \{s_1, \dots, s_k\}$ i macierzy przejść P zaczyna się w stanie s_i ($X_0 = s_i$), to *czas pierwszych odwiedzin* dla stanu s_j definiujemy jako

$$T_{i,j} = \min\{n \geq 1 : X_n = s_j\}.$$

Przyjmujemy przy tym, że jeśli łańcuch nigdy nie dochodzi do stanu s_j , to $T_{i,j} = \infty$.

Definicja 6.28. Średnim czasem odwiedzin nazywamy wartość:

$$\tau_{i,j} = \mathbb{E}(| T_{i,j} |).$$

W przypadku, gdy $i = j$ wartość $\tau_{i,i}$ nazywamy *średnim czasem powrotu* dla stanu s_i .

Możemy teraz przedstawić lemat, który wykorzystany będzie w kluczowym twierdzeniu tego rozdziału tj. Twierdzeniu Ergodycznym o zbieżności do rozkładu stacjonarnego 6.30.

Lemat 6.29:

Dla każdego nieredukowalnego i nieokresowego łańcucha Markowa o skończonym zbiorze stanów $S = \{s_1, \dots, s_k\}$ i macierzy przejść P mamy, że dla każdych dwóch stanów $s_i, s_j \in S$, jeśli łańcuch zaczyna się w s_i , to

$$\mathbb{P}(T_{i,j} < \infty) = 1.$$

Co więcej, średni czas powrotu $\tau_{i,j}$ jest skończony

$$\mathbb{E}(| T_{i,j} | < \infty).$$

Twierdzenie 6.30 (Twierdzenie Ergodyczne o zbieżności łańcuchów Markowa).

Niech (X_1, X_2, \dots) będzie nieredukowalnym i nieokresowym łańcuchem Markowa o skończonym zbiorze stanów $S = \{s_1, \dots, s_k\}$, macierzy przejść P oraz rozkładzie początkowym $\mu^{(0)}$. Istnieje wtedy unikalny rozkład stacjonarny π , taki, że niezależnie od rozkładu początkowego $\mu^{(0)}$ zachodzi $\mu^{(n)} \rightarrow \pi$, gdy $(n \rightarrow \infty)$.

7. Wykład 7: Rozkłady stacjonarne. Coupling.

Wstęp

W tym wykładzie pokażemy istnienie rozkładu stacjonarnego dla dowolnego nierozkładalnego i nieokresowego łańcucha Markowa oraz jego postać. Korzystając ze zbieżności wykazemy unikalność owego rozkładu (tzn. że jest on wyznaczony jednoznacznie). Przedstawimy również metodę Coupling'u, która umożliwi wykazanie zbieżności do rozkładu stacjonarnego (stabilizowanie się łańcucha).

Przypomnijmy, że przestrzeń stanów $S = \{s_1, s_2, \dots, s_k\}$, $T_{i,j}$ i $\tau_{i,j}$ będą takie jak w poprzednim rozdziale. Udowodnimy teraz lemat 6.29.

Dowód lematu 6.29. Z Wniosku 6.24 wiemy, że istnieje M takie, że macierz przejść P^M składa się z dodatnich wartości, czyli innymi słowy istnieje takie M , że poczynawszy od momentu M z dodatnim prawdopodobieństwem osiągniemy z każdego stanu s_i dowolny stan s_j . Oznaczmy przez $\alpha = \min_{i,j} \{(P^M)_{i,j}\} > 0$. Wówczas prawdopodobieństwo nie osiągnięcia stanu s_j w mniej niż M krokach wynosi

$$\mathbb{P}(T_{i,j} > M \mid X_0 = s_i) \leq \mathbb{P}(X_M \neq s_j \mid X_0 = s_i) \leq 1 - \alpha.$$

W dalszym ciągu pomocnym będzie poniższe szacowanie prawdopodobieństwa warunkowego, że stan s_j nie zostanie osiągnięty w mniej niż $2M$ krokach, gdy w mniej niż M krokach nie wystąpił ten stan:

$$\begin{aligned} \mathbb{P}(T_{i,j} > 2M \mid T_{i,j} > M, X_0 = s_i) &\leq \mathbb{P}(X_{2M} \neq s_j \mid T_{i,j} > M, X_0 = s_i) = \\ &= \sum_{\substack{l=1 \\ l \neq j}}^k \mathbb{P}(X_{2M} \neq s_j \mid T_{i,j} > M, X_M = s_l, X_0 = s_i) = \\ &= \sum_{\substack{l=1 \\ l \neq j}}^k \mathbb{P}(X_{2M} \neq s_j \mid T_{i,j} > M, X_M = s_l) = \\ &= \sum_{\substack{l=1 \\ l \neq j}}^k \mathbb{P}(X_{2M} \neq s_j \mid X_M = s_l) \mathbb{P}(X_M = s_l \mid T_{i,j} > M) \leq \\ &\leq (1 - \alpha) \sum_{\substack{l=1 \\ l \neq j}}^n \mathbb{P}(X_M = s_l \mid T_{i,j} > M) \leq \\ &\leq 1 - \alpha. \end{aligned}$$

Oszacujmy teraz prawdopodobieństwo, że stan s_j nie zostanie osiągnięty w mniej niż $2M$ krokach:

$$\begin{aligned}\mathbb{P}(T_{i,j} > 2M \mid X_0 = s_i) &\stackrel{\star}{=} \mathbb{P}(T_{i,j} > M \mid X_0 = s_i) \mathbb{P}(T_{i,j} > 2M \mid T_{i,j} > M, X_0 = s_i) \leq \\ &\leq (1 - \alpha)(1 - \alpha) = (1 - \alpha)^2\end{aligned}$$

gdzie przejście \star ze wzoru łańcuchowego.

Powtarzając analogiczne rozumowanie dla prawdopodobieństw

$$\mathbb{P}(T_{i,j} > 3M \mid T_{i,j} > M, X_0 = s_i), \mathbb{P}(T_{i,j} > 4M \mid T_{i,j} > M, X_0 = s_i) \text{ itd.}$$

otrzymujemy, że dla dowolnego $l \in \mathbb{N}$

$$\mathbb{P}(T_{i,j} > lM \mid X_0 = s_i) \leq (1 - \alpha)^l,$$

zatem prawdopodobieństwo, że stan s_j nie zostanie nigdy osiągnięty wynosi

$$\begin{aligned}\mathbb{P}(T_{i,j} = \infty \mid X_0 = s_i) &= \mathbb{P}\left(\bigcap_{l=1}^{\infty} \{T_{i,j} > lM \mid X_0 = s_i\}\right) = \lim_{l \rightarrow \infty} \mathbb{P}(T_{i,j} > lM \mid X_0 = s_i) \leq \\ &\leq \lim_{l \rightarrow \infty} (1 - \alpha)^l = 0.\end{aligned}$$

Na zakończenie obliczmy oczekiwaną liczbę kroków potrzebną do osiągnięcia stanu s_j

$$\begin{aligned}\tau_{i,j} &\stackrel{\star\star}{=} \sum_{n=1}^{\infty} \mathbb{P}(T_{i,j} > n) = \sum_{l=0}^{\infty} \sum_{n=lM}^{(l+1)M-1} \mathbb{P}(T_{i,j} > n) \leq \sum_{l=0}^{\infty} \sum_{n=lM}^{(l+1)M-1} \mathbb{P}(T_{i,j} > lM) = \\ &= M \sum_{l=0}^{\infty} \mathbb{P}(T_{i,j} > lM) \leq M \sum_{l=0}^{\infty} (1 - \alpha)^l = \frac{M}{\alpha} = \text{const}.\end{aligned}$$

Równość $\star\star$ wynika z Twierdzenia 11. z rozdziału §5.6 z Wstęp do teorii prawdopodobieństwa J.Jakubowski, R.Sztencel. \square

Dowód istnienia rozkładu stacjonarnego w Twierdzeniu Ergodycznym 6.30. Dla uproszczenia wprowadźmy następujące oznaczenia:

$$\begin{aligned}i &= s_i, \\ T_l &= T_{l,l}.\end{aligned}$$

(i – stan s_i naszego łańcucha, T_l – czas pierwszego powrotu z l do l). Ustalmy dowolny stan $l \in [k]$. W dalszej części wszystkie zdarzenia są rozpatrywane w przestrzeni warunkowej z warunkiem $X_0 = s_l$. Dla każdego $i \in [k]$ będącego pewnym stanem definiujemy zmienną losową

$$N_i(l) = \sum_{n=1}^{\infty} I_{\{X_n=i\} \cap \{T_l \geq n\}}.$$

Jeśli $i \neq l$ to powyższa zmienna losowa liczy ile razy odwiedzony zostanie stan i po drodze do l . Oznaczmy przez $\rho_i(l) = \mathbb{E}(N_i(l) \mid X_0 = l)$

Uwaga, oczywiście $N_l(l) = 1$, czyli również $\rho_i(i) = 1$. Zauważmy, że:

$$T_l = \sum_{i=1}^k N_i(l),$$

gdyż wszystkie kroki jakie pokonamy idąc z l do l możemy rozbić ze względu na liczbę odwiedzin w każdym ze stanów. Zatem

$$\mu_l = \mathbb{E}(T_l) = \sum_{i=1}^k \rho_i(l),$$

i na mocy lematu 6.29 $\mu_l < \infty$, czyli dla każdego $i \in [k]$ mamy $\rho_i(l) < \infty$.

Kandydatem na rozkład dla każdego l jest

$$\boldsymbol{\pi}(l) = \frac{\boldsymbol{\rho}(l)}{\mu_l},$$

gdzie $\boldsymbol{\rho}(l) = (\rho_1(l), \dots, \rho_k(l))$, czyli $\boldsymbol{\pi}(l) = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_k) = \left(\frac{\rho_1(l)}{\mu_l}, \dots, \frac{\rho_k(l)}{\mu_l}\right)$. Pokażemy, że jest to rozkład stacjonarny, czyli spełnia Definicję 6.25.

Mamy $N_i(l) \geq 0$, więc $\rho_i(l) \geq 0$, a stąd $\boldsymbol{\pi}_i(l) \geq 0$ oraz

$$\sum_{i=1}^k \boldsymbol{\pi}_i(l) = \sum_{i=1}^k \frac{\rho_i(l)}{\mu_l} = \frac{1}{\mu_l} \sum_{i=1}^k \rho_i(l) = 1$$

Zatem $\boldsymbol{\pi}(l)$ jest rozkładem prawdopodobieństwa, czyli spełnia punkt (i) definicji 6.25. Sprawdźmy teraz prawdziwość punktu (ii) tej definicji:

$$\begin{aligned} \rho_i(l) &= \sum_{n=1}^{\infty} \mathbb{P}(X_n = i, T_l \geq n \mid X_0 = l) = \\ &= p_{li} + \sum_{n \geq 2} \sum_{\substack{j=1 \\ j \neq l}}^k \mathbb{P}(X_n = i, T_l \geq n, X_{n-1} = j \mid X_0 = l) = \\ &= p_{li} + \sum_{n \geq 2} \sum_{\substack{j=1 \\ j \neq l}}^k \mathbb{P}(X_{n-1} = j, T_l \geq n) \mathbb{P}(X_n = i \mid T_l \geq n, X_{n-1} = j) = \\ &= \rho_l p_{li} + \sum_{n \geq 2} \sum_{\substack{j=1 \\ j \neq l}}^k \mathbb{P}(X_{n-1} = j, T_l \geq n-1) \mathbb{P}(X_n = i \mid X_{n-1} = j) = \\ &= \rho_l(l) p_{li} + \sum_{\substack{j=1 \\ j \neq l}}^k p_{ji} \sum_{n=2}^{\infty} \mathbb{P}(X_{n-1} = j, T_l \geq n-1) = \\ &= \rho_l(l) p_{li} + \sum_{\substack{j=1 \\ j \neq l}}^k p_{ji} \rho_j(l) = \\ &= \sum_{j=1}^k \rho_j(l) p_{ji}, \end{aligned}$$

gdzie na końcu można łatwo dostrzec mnożenie macierzy, a stąd :

$$\begin{aligned}\rho(l) &= \rho(l)P / \cdot \frac{1}{\mu_l} \\ \pi(l) &= \pi(l)P\end{aligned}$$

Co dowodzi, że $\pi(l)$ jest rozkładem stacjonarnym zgodnie z definicją 6.25.
c.d.n. □

7.1. Postać rozkładu stacjonarnego

Wróćmy do pozostałych tez Twierdzenia Ergodycznego 6.30. Będziemy starali się pokazać, że zachodzi następujący ciąg implikacji:

$$\text{zbieżność} \Rightarrow \text{jedyność} \Rightarrow \frac{\rho_i(l)}{\mu_l} = \frac{\rho_i(i)}{\mu_i} = \frac{1}{\mu_i} \Rightarrow \pi = \left(\frac{1}{\mu_1}, \dots, \frac{1}{\mu_k} \right) - \text{rozkład stacjonarny.}$$

Na potrzeby dalszej części tego rozdziału założymy prawdziwość następującego twierdzenia mówiącego, że rozkład brzegowy zbiega do rozkładu stacjonarnego.

Twierdzenie 7.1.

Jeśli π jest rozkładem stacjonarnym, to dla każdego $\mu^{(0)}$ mamy

$$\mu^{(n)} \xrightarrow{TV} \pi \Leftrightarrow \lim_{n \rightarrow \infty} d_{TV}(\mu^{(n)}, \pi) = 0$$

Korzystając z Twierdzenia 7.1 można udowodnić kawałek Twierdzenia Ergodycznego 6.30, mianowicie jedyność rozkładu stacjonarnego.

Dowód jednoznaczności wyznaczenia rozkładu stacjonarnego w Twierdzeniu Ergodycznym 6.30. Przypuśćmy, że istnieją dwa rozkłady stacjonarne π i π' . Niech $\mu^{(0)} = \pi'$. Wtedy $\mu^{(n)} = \pi'$ dla każdego $n \in \mathbb{N}$, ponieważ jest on rozkładem stacjonarnym. Ponieważ jednak π jest również rozkładem stacjonarnym, więc z Twierdzenia 7.1 otrzymujemy

$$\mu^{(0)} \xrightarrow{TV} \pi \Leftrightarrow \lim_{n \rightarrow \infty} d_{TV}(\pi', \pi) = 0,$$

a zatem $d_{TV}(\pi', \pi) = 0 \Rightarrow \pi = \pi'$ zgodnie z Definicją 6.26.
c.d.n. □

Zauważmy, że jedyną brakującą częścią dowodu Twierdzenia Ergodycznego 6.30 jest udowodnienie zbieżności, czyli Twierdzenia 7.1. W celu jej udowodnienia i ostatecznie Twierdzenia Ergodycznego 6.30 wprowadźmy pojęcie couplingu.

7.2. Coupling

Definicja 7.2. Załóżmy, że dane są dwie zmienne losowe X, Y na przestrzeniach probabilistycznych odpowiednio Ω_1, Ω_2 . Konstruujemy „kopie” (*coupling*) X', Y' zmiennych X i Y na tej samej przestrzeni Ω tak, aby ich rozkłady brzegowe były takie same jak rozkłady X i Y oraz były one ze sobą powiązane.

W przypadku couplingu przestrzenie produktowe są praktycznie bezużyteczne.

Definicja 7.3. Mówimy, że X dominuje Y , gdy dla każdej wartości dystrybuant x spełniona jest nierówność $F_X(x) \leq F_Y(x)$, czyli $\mathbb{P}(X \leq x) \leq \mathbb{P}(Y \leq x)$. (Jest tak np. gdy $X \geq Y$.) W dalszej części będziemy wykorzystywać oznaczenie $X \geq_{st} Y$.

Twierdzenie 7.4.

Jeśli $X \geq_{st} Y$ to istnieje (Ω, \mathcal{F}, P) oraz X' i Y' na (Ω, \mathcal{F}, P) takie, że $F_{X'} = F_X$, $F_{Y'} = F_Y$ oraz $\mathbb{P}(X' \geq Y') = 1$.

Dowód. Niech $\Omega = \langle 0, 1 \rangle$, $\mathcal{F} = \mathcal{B}_{\mathbb{R}}(\langle 0, 1 \rangle)$ oraz $P = \lambda_1(\langle 0, 1 \rangle)$. Dla dowolnej dystrybuanty F zdefiniujemy zmienną losową Z_F na Ω w następujący sposób:

$$Z_F(\omega) = \inf \{z: \omega \leq F(z)\}$$

Zauważmy, że $\omega \leq F(z) \Leftrightarrow Z_F(\omega) \leq z$. Stąd:

$$F_{Z_F}(z) = \mathbb{P}(Z_F \leq z) = \mathbb{P}(\langle 0, F(z) \rangle) = \lambda_1(\langle 0, F(z) \rangle) = F(z)$$

Niech $X' = Z_{F_X}$, $Y' = Z_{F_Y}$. Wówczas $X' \geq Y'$ prawie wszędzie, czyli $\mathbb{P}(X' \geq Y') = 1$, co kończy dowód. \square

Wiemy, że $Bin(n, \frac{\lambda}{n}) \rightarrow Po(\lambda)$, gdy $n \rightarrow \infty$ (zbieżność szeregów liczbowych dla $k \in \{0, \dots, n\}$). Jak dobre jest to przybliżenie?

Twierdzenie 7.5 (Cam, 1960).

Jeśli X_1, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładzie Bernoulliego, gdzie $\mathbb{E}(X_i) = p_i$ oraz liczba sukcesów

$$S = \sum_{r=1}^n X_r$$

to

$$d_{TV}(S, P) \leq \sum_{r=1}^n p_r^2$$

gdzie

$$P \sim Po\left(\lambda = \sum_{r=1}^n p_r\right).$$

W szczególności jeżeli $p_r = \frac{\lambda}{n}$, to $d_{TV}(S, P) \leq \frac{\lambda^2}{n}$.

8. Wykład 8: Zbieżność poissonowska. Twierdzenie ergodyczne. Łańcuchy odwracalne.

Wstęp

W pierwszej części dokończymy zagadnienia z poprzedniego wykładu ukazując zastosowania couplingu na kilku przykładach. Następnie dokończymy dowód Twierdzenie Ergodyczne 6.30. W drugiej części wprowadzimy odwracalne łańcuchy Markowa i podamy kilka prostych przykładów.

8.1. Dokończenie poprzedniego wykładu

Do udowodnienia Twierdzenia 7.5 potrzebować będziemy następującego lematu:

Lemat 8.1:

Dla każdego $x \in \langle 0, 1 \rangle$, zachodzi nierówność $e^{-x} \geq 1 - x$.

Dowód. Niech $\varphi(x) = e^{-x} - 1 + x$. Jak łatwo można zauważyć, jest to funkcja różniczkowalna na przedziale $\langle 0, 1 \rangle$. Otrzymujemy $\varphi'(x) = -e^{-x} + 1 \leq 0$, gdyż $\forall x \in \langle 0, 1 \rangle e^{-x} \geq 1$. Jest to zatem funkcja malejąca na przedziale $\langle 0, 1 \rangle$. Czyli $\varphi(x)$ przyjmuje najmniejszą wartość w dziedzinie w punkcie 1.

$$\forall_{x \in \langle 0, 1 \rangle} \varphi(x) \geq \varphi(1) = e^{-1} \geq 0.$$

Czyli $e^{-x} - 1 + x \geq 0$ □

Dowód Twierdzenia 7.5. Budujemy coupling niezależnych zmiennych losowych (X_r, Y_r) na przestrzeni $\{0, 1\} \times \{0, 1, 2, \dots\}$

$$\mathbb{P}(X_r = x, Y_r = y) = \begin{cases} 1 - p_r & x = y = 0 \\ e^{-p_r} - 1 + p_r & x = 1, y = 0 \\ \frac{p_r^y e^{-p_r}}{y!} & x = 1, y > 0 \\ 0 & \text{w p.p.} \end{cases}$$

Patrząc na postać funkcji prawdopodobieństwa, łatwo można zauważyć, że X_r ma rozkład $Ber(p_r)$ oraz Y_r ma rozkład $Po(p_r)$. Postaramy się ukazać ideę wprowadzenia takiego couplingu tym, którzy nie są w stanie dojrzeć jej już za pierwszym razem. Z postaci funkcji oczywiście natychmiast wynika, że

$$\mathbb{P}(X_r \neq Y_r) = p_r - p_r e^{-p_r}.$$

W oparciu o Lemat 8.1 otrzymujemy

$$\mathbb{P}(X_r \neq Y_r) = p_r - p_r e^{-p_r} = p_r(1 - e^{-p_r}) \leq p_r p_r = p_r^2.$$

Wprowadźmy definicje:

$$S = \sum_{r=1}^n X_r, \quad P = \sum_{r=1}^n Y_r$$

tak, aby powstały zmienne losowe postaci podanej w treści twierdzenia. Zauważmy, że

$$\begin{aligned} |\mathbb{P}(S = k) - \mathbb{P}(P = k)| &= |\mathbb{P}(S = k, P \neq k) - \mathbb{P}(S \neq k, P = k)| \leq \\ &\leq \mathbb{P}(S = k, P \neq k) + \mathbb{P}(S \neq k, P = k) \end{aligned}$$

gdzie ostatnia nierówność wynika z faktu, że prawdopodobieństwo jest nieujemne. Wróćmy na główny tor. Z definicji 6.26 mamy, że:

$$\begin{aligned} d_{TV}(S, P) &= \frac{1}{2} \sum_k |\mathbb{P}(S = k) - \mathbb{P}(P = k)| \leq \frac{1}{2} \sum_k (\mathbb{P}(S = k, P \neq k) + \mathbb{P}(S \neq k, P = k)) = \\ &= 2 \frac{1}{2} \mathbb{P}(S \neq P) = \mathbb{P}(S \neq P) \leq \mathbb{P}(\exists_r, X_r \neq Y_r) \leq \\ &\leq \sum_{r=1}^n \mathbb{P}(X_r \neq Y_r) \leq \sum_{r=1}^n p_r^2 \end{aligned}$$

□

Czas powrócić do potężnego Twierdzenia Ergodycznego 6.30.

Dokończenie dowodu Twierdzenia 6.30, dowód Twierdzenia 7.1. Aby ostatecznie udowodnić Twierdzenie 6.30 pozostaje wykazać zbieżność $\mu^{(n)} \xrightarrow{TV} \boldsymbol{\pi}$, czyli udowodnić Twierdzenie 7.1. Niech $S = \{1, 2, \dots, k\}$ będzie zbiorem stanów, a $\boldsymbol{\pi}$ rozkładem stacjonarnym rozpatrywanego łańcucha Markowa $X = (X_n)$. Z Definicji 6.26 (po usunięciu nic nie znaczącej w tym kontekście $\frac{1}{2}$) musimy pokazać, że

$$\lim_{n \rightarrow \infty} \sum_{i=1}^k |\mu_i^{(n)} - \pi_i| = 0$$

Zapusemy drugi łańcuch $Y = (Y_n)$ mający tą samą macierz przejść, ale inne warunki początkowe: X ma na początku rozkład $\mu^{(0)}$, natomiast Y ma $\boldsymbol{\pi}$.

Aby nie zagmatwać się w labiryncie trików przedstawmy główną ideę tego dowodu. Jeśli $X_m = Y_m$ dla pewnego m , to dla każdego $n \geq m$ X_n i Y_n mają ten sam rozkład, gdyż jest to rozkład stacjonarny.

Wprowadźmy zmienną losową $T = \min\{m : X_m = Y_m\}$, czyli pierwszy moment czasu, w którym zmienne X i Y się spotkają. Wówczas dla każdego $i \in S$ oraz dla każdego $n \in \mathbb{N}$ mamy:

$$\mu_i^{(n)} = \mathbb{P}(X_i = i) = \mathbb{P}(X_n = i | T \leq n) \mathbb{P}(T \leq n) + \mathbb{P}(X_n = i | T > n) \mathbb{P}(T > n)$$

Zauważmy, że w wyrażeniu $\mathbb{P}(X_n = i | T \leq n)$ możemy wymienić X_n na Y_n , gdyż w tej przestrzeni warunkowej z definicji T są one stakie same. Wówczas

$$\begin{aligned}\mu_i^{(n)} &= \mathbb{P}(Y_n = i | T \leq n) \mathbb{P}(T \leq n) + \mathbb{P}(X_n = i | T > n) \mathbb{P}(T > n) \leq \\ &\leq \mathbb{P}(Y_n = i) + \mathbb{P}(T > n)\end{aligned}$$

gdź prawdopodobieństwo jest funkcją mniejszą od 1. Stąd

$$\mu_i^{(n)} \leq \pi_i + \mathbb{P}(T > n)$$

Przez symetrię dostaniemy

$$\pi_i \leq \mu_i^{(n)} + \mathbb{P}(T > n)$$

co implikuje

$$|\mu_i^{(n)} - \pi_i| \leq \mathbb{P}(T > n).$$

Jeśli więc pokażemy, że $\mathbb{P}(T > n)$ dąży do 0 z n biegnącym do nieskończoności, udowodnimy interesującą nas zbieżność, gdyż skończona suma składników zbiegających do 0 oczywiście również zbiega do 0. Ponadto z Wniosku 6.24 wiemy, że

$$\exists M : \alpha = \min_{i,j} (P^M)_{i,j} > 0$$

Wtedy

$$\begin{aligned}\mathbb{P}(T \leq M) &\geq \mathbb{P}(X_M = 1, Y_M = 1) \geq \alpha^2 \\ \mathbb{P}(T > M) &\leq 1 - \alpha^2 \implies \mathbb{P}(T > lM) \leq (1 - \alpha^2)^l\end{aligned}$$

Wybierając $l = \frac{n}{M}$ dostaniemy:

$$\mathbb{P}(T > n) \leq \mathbb{P}(T > lM) \leq (1 - \alpha^2)^{\frac{n}{M}} \rightarrow 0$$

przy $n \rightarrow \infty$, bo $(1 - \alpha^2) < 1$. □

Uwaga 8.2. Gdy $\mu^{(0)} = (0, \dots, 0, 1, 0, \dots, 0)$ (jedynek jest na i -tym miejscu), wówczas również $\mu^{(n)} \rightarrow \pi$. Innymi słowy, nie jest istotne skąd zaczynamy, ale dokąd zmierzamy. Po pewnej liczbie kroków stan początkowy „rozplywa się” i od pewnego momentu nie ważne jest już skąd zaczynaliśmy.

8.2. Odwracalne Łańcuchy Markowa

W tej części wykładu wprowadzone zostaną odwracalne łańcuchy Markowa, ich własności, zastosowania oraz przykłady.

Definicja 8.3. Rozkład π nazywamy *odwracalnym*, gdy

$$\forall_{i,j \in S} \pi_i p_{ij} = \pi_j p_{ji}$$

gdzie π_i jest i -tą współrzędną wektora prawdopodobieństwa rozkładu π , a p_{ij} prawdopodobieństwem przejścia ze stanu s_i do stanu s_j .

Definicja 8.4. Łańcuch Markowa nazywamy *odwracalnym* gdy jego rozkład jest odwracalny.

Lemat 8.5:

Rozkład odwracalny π jest rozkładem stacjonarnym.

Dowód.

$$\pi_j = \pi_j \cdot 1 = \pi_j \sum_{i=1}^k p_{ji} = \sum_{i=1}^k \pi_j p_{ji} = \sum_{i=1}^k \pi_i p_{ij}.$$

□

Lemat 8.6:

Niech X, Y będą zmiennymi losowymi takimi, że dla każdego $n \in \{0, 1, \dots, N\}$, $Y_n = X_{N-n}$ oraz niech $\mu^{(0)} = \pi$ będzie rozkładem odwracalnym. Wtedy

$$\mathbb{P}(Y_{n+1} = j | Y_n = i) = \frac{\pi_j}{\pi_i} p_{ji} \quad (= p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i))$$

Dowód. Pokażemy, że Y jest łańcuchem Markowa.

$$\begin{aligned} \mathbb{P}(Y_{n+1} = i_{n+1} | Y_n = i_n, \dots, Y_0 = i_0) &\stackrel{\text{pw}}{=} \frac{\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n, Y_{n+1} = i_{n+1})}{\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n)} \stackrel{X_{N-n}=Y_n}{=} \\ &= \frac{\mathbb{P}(X_{N-n-1} = i_{n+1}, \dots, X_N = i_0)}{\mathbb{P}(X_{N-n} = i_n, \dots, X_N = i_0)} \stackrel{\text{wł. 8.3}}{=} \\ &= \frac{\pi_{i_{n+1}} p_{i_{n+1}, i_n} \cdot p_{i_n, i_{n-1}} \cdot \dots \cdot p_{i_1, i_0}}{\pi_{i_n} \cdot p_{i_n, i_{n-1}} \cdot \dots \cdot p_{i_1, i_0}} = \\ &= \frac{\pi_{i_{n+1}} \cdot p_{i_{n+1}, i_n}}{\pi_{i_n}} = \frac{\pi_{i_n} \cdot p_{i_n, i_{n+1}}}{\pi_{i_n}} = p_{i_n, i_{n+1}} = \\ &= \mathbb{P}(X_{N-n-1} = i_{n+1} | X_{N-n} = i_n) \stackrel{X_{N-n-1}=Y_{n+1}}{=} \\ &= \mathbb{P}(Y_{n+1} = i_{n+1} | Y_n = i_n) \end{aligned}$$

„pw” oznacza prawdopodobieństwo warunkowe, a „wł” wzór łańcuchowy. □

Przykład 8.7 (Spacer po grafie). Niech $G = (V, E)$ będzie grafem prostym. Prawdopodobieństwo przejścia z wierzchołka i do wierzchołka j wynosi

$$p_{ij} = \begin{cases} \frac{1}{d_i} & \text{gdy } \{i, j\} \in E \\ 0 & \text{gdy } \{i, j\} \notin E \end{cases}$$

gdzie d_i oznacza stopień wierzchołka i . Oznaczmy przez $e = |E|$. Wówczas tak zdefiniowany łańcuch jest odwracalnym łańcuchem Markowa, gdyż rozkład $\pi = \left(\frac{d_1}{2e}, \dots, \frac{d_n}{2e}\right)$ jest rozkładem stacjonarnym i odwracalnym:

$$\begin{aligned} \pi_i p_{ij} &= \frac{d_i}{2e} \frac{1}{d_i} = \frac{1}{2e} \\ \pi_j p_{ji} &= \frac{d_j}{2e} \frac{1}{d_j} = \frac{1}{2e} \end{aligned}$$

Przykład 8.8 (Proces narodzin i śmierci). Niech $S = [k] = \{1, 2, 3, \dots, k\}$,

$$p_{ij} = \begin{cases} > 0 & \text{gdy } |i - j| = 1 \\ 0 & \text{gdy } |i - j| \geq 2 \end{cases}$$

tj. prawdopodobieństwo przejścia ze stanu i do stanu j wynosi zero gdy stany te spełniają nierówność $|i - j| \geq 2$, zaś w przeciwnym przypadku prawdopodobieństwo to jest dodatnie. Dla $j = i$ prawdopodobieństwo P_{ii} jest dowolne.

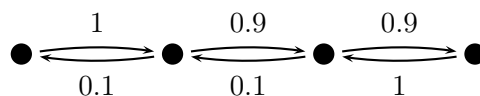
Otrzymujemy macierz przejść postaci

$$\begin{pmatrix} ? & > 0 & 0 & \dots & 0 \\ > 0 & ? & > 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots \\ 0 & \dots & > 0 & ? & > 0 \\ 0 & \dots & 0 & > 0 & ? \end{pmatrix}$$

Dla $k = 4$ macierz przejść może być następująca:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.1 & 0 & 0.9 & 0 \\ 0 & 0.1 & 0 & 0.9 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

co daje graf przejść:



Niech teraz π_1^* będzie dowolne, a pozostałe zdefiniujemy jako

$$\pi_i^* = \frac{p_{i-1,i}}{p_{i,i-1}} \pi_{i-1}^*.$$

Rozkład π dany wzorem

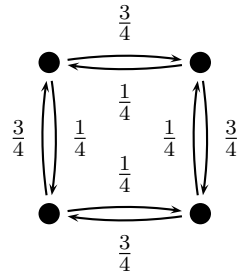
$$\pi_i = \frac{\pi_i^*}{\sum_{j=1}^k \pi_j^*}$$

jest rozkładem stacjonarnym. Łatwo sprawdzić, że w naszym konkretnym przykładzie rozkład ten przyjmie następującą postać:

$$\pi = \left(\frac{1}{182}, \frac{10}{182}, \frac{90}{182}, \frac{81}{182} \right)$$

Przykład 8.9. Przedstawimy teraz przykład łańcucha z rozkładem stacjonarnym i nieodwracalnym.

Niech dany będzie łańcuch Markowa taki jak na rysunku:



Rozkład $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ jest stacjonarny. Nie jest on odwracalny, ponieważ spacerując po tym łańcuchu mamy „tendencję” do „przemieszczania się” zgodnie z kierunkiem wskazówek zegara. Nie ma symetryczności.

9. Wykład 9: Markov Chain Monte Carlo

Wstęp

W tej części wykładu zajmiemy się zastosowaniami łańcuchów Markowa w informatyce. Naszym celem będzie przedstawienie na kilku przykładach tzw. metody MCMC (ang. Markov Chain Monte Carlo), czyli metody Monte Carlo z wykorzystaniem łańcuchów Markowa.

Ogólnie metoda MCMC jest zaliczana do klasy metod symulacyjnych, gdyż stosowana jest do matematycznego modelowania procesów zbyt złożonych, aby można było przewidzieć ich wyniki za pomocą prostego modelu zastępczego, który można zaimplementować na komputerze, a następnie przeprowadzić symulacje tegoż modelu. Istotną rolę w metodzie MCMC odgrywa losowanie wielkości charakteryzujących proces, przy czym losowanie dokonywane jest zgodnie ze znanym nam rozkładem.

Metody MCMC powstały w odpowiedzi na stawiane problemy przy projektowaniu eksperymentów fizycznych (w fizyce statystycznej czy chemii fizycznej), stąd są one bardzo rozpowszechnione w tych naukach.

Warto wspomnieć, że twórcami metody Monte Carlo była grupa wybitnych matematyków – Nicholas Metropolis, John von Neumann oraz Stanisław Ulam, pracujących w ośrodku badań jądrowych w Los Alamos w latach 40-tych XX wieku.

9.1. Idea MCMC

Niech Ω będzie skończoną przestrzenią probabilistyczną o rozbudowanej strukturze ze skomplikowanym rozkładem π . Metoda MCMC polega na konstrukcji nieredukowalnego nieokresowego łańcucha Markowa o danej przestrzeni stanów $S = \Omega$, macierzy przejść P o nieskomplikowanej budowie oraz rozkładzie stacjonarnym π . Celem będzie skorzystanie z Twierdzenia Ergodycznego 6.30, które gwarantuje, że obrany rozkład początkowy dla takiego łańcucha po wystarczająco długim czasie dąży do rozkładu stacjonarnego.

9.2. Przykłady

Przykład 9.1 (The hard-core model).

Niech dany będzie graf $G = (V, E)$, gdzie $|V| = k$ jest jego liczbą wierzchołków oraz $|E| = l$ liczbą krawędzi. Wierzchołki grafu są ponumerowane w sposób losowy wartościami 0 oraz 1. Oczywiście każde takie przyporządkowanie wartości 0, 1 jest funkcją, tutaj zwaną *konfiguracją*, jedną z $\{0, 1\}^V$ możliwych. *Konfiguracją dopuszczalną* nazywamy taką, w

której żadne dwie „1” nie są połączone krawędzią (w teorii grafów powiemy, że zbiór jedynek w tej konfiguracji jest tzw. zbiorem niezależnym).

Niech Z_G będzie liczbą dopuszczalnych konfiguracji w grafie G . Dla dowolnej konfiguracji $\xi \in \{0, 1\}^V$ bierzemy miarę probabilistyczną

$$\mu_G(\xi) = \begin{cases} \frac{1}{Z_G} & \text{jeżeli } \xi \text{ jest dopuszczalna} \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Oznacza to, że losowanie dowolnej konfiguracji sprowadza się do wybrania każdej z dopuszczalnych z równym prawdopodobieństwem.

Naturalnym pytaniem, które się tu pojawia jest: *Jaka jest średnia liczba „1” w losowej konfiguracji wybranej zgodnie z rozkładem $\mu_G(\xi)$?*

Niech $n(\xi)$ będzie liczbą „1” w ξ (w teorii grafów jest to rozmiar zbioru niezależnego), a X losową konfiguracją wybraną zgodnie z $\mu_G(\xi)$. Wówczas średnia wielkość konfiguracji dopuszczalnej w G (średni rozmiar zbioru niezależnego) jest równa

$$\mathbb{E}(n(X)) = \sum_{\xi \in \{0,1\}^V} n(\xi) \cdot \mu_G(\xi) = \frac{1}{Z_G} \sum_{\xi \in \{0,1\}^V} n(\xi) \cdot I_{\{\xi \text{ jest dopuszczalna}\}}$$

Zauważmy, że powyższa suma zawiera bardzo dużo składników, co znacznie zniechęca do jej obliczenia. A gdyby do postawionego problemu użyć symulacji? Pojawia się jednak problem: *Ile wykonać symulacji, aby otrzymać satysfakcjonujący wynik nieodbiegający zbyt od powyższego, oraz w jaki sposób wygenerować losowe ξ będące dopuszczalnym?*

Pomysł polega na konstrukcji nieredukowalnego i nieokresowego łańcucha Markowa $X = (X_0, X_1, \dots)$, na przestrzeni stanów $S = \{\xi \in \{0, 1\}^V : \xi \text{ dopuszczalna}\}$, mającego macierz przejść $p_{\xi, \xi'} = \mathbb{P}(X_{n+1} = \xi' | X_n = \xi)$ oraz rozkład stacjonarny μ_G .

Łańcuch Markowa spełniający powyższe wymagania tworzymy na podstawie następującego algorytmu:

Algorytm 9.2.

1. Wylosuj $v \in V$ (z równym prawdopodobieństwem)
2. $W :=$ wynik rzutu symetryczną monetą
3. Jeśli $W =$ „Orzeł” oraz $\forall w \in N_G(v) : X_n(w) = 0$, to $X_{n+1}(v) = 1$;
W przeciwnym przypadku $X_{n+1}(v) = 0$
4. $\forall w \neq v : X_{n+1}(w) = X_n(w)$

Powstały w ten sposób łańcuch Markowa jest nieredukowalny i nieokresowy (zadanie domowe 9.1.). Pozostaje jeszcze pokazać, że μ_G jest jego rozkładem stacjonarnym, co sprowadza się do sprawdzenia jego odwracalności (zgodnie z Lematem 8.5). Wiemy, że μ_G będzie odwracalny, gdy dla każdego dopuszczalnych ξ i ξ' $\mu_G(\xi) \cdot p_{\xi, \xi'} = \mu_G(\xi') \cdot p_{\xi', \xi}$. Niech $d = d(\xi, \xi')$ będzie odległością Hamminga, czyli liczbą wierzchołków różniących się wartościami w ξ i ξ' . Może ona przyjmować następujące wartości:

- $d = 0$, gdy $\xi = \xi'$ (konfiguracje są identyczne).
- $d = 1$, gdy istnieje dokładnie jeden wierzchołek v dla którego $\xi(v) = 1$ oraz $\xi'(v) = 0$ (lub na odwrót) (konfiguracje różnią się na dokładnie jednej pozycji).

— $d \geq 2$, (nie można w jednym kroku przejść z jednej konfiguracji do drugiej, więc $p_{\xi', \xi} = 0$)
Otrzymujemy zatem: $\mu_G(\xi) \cdot p_{\xi, \xi'} = \frac{1}{Z_G} \frac{1}{2k} = \mu_G(\xi') \cdot p_{\xi', \xi}$, co na mocy Lematu 8.5 oznacza, że μ_G jest żądanym rozkładem stacjonarnym.

Podany powyżej algorytm jest szczególnym przypadkiem algorytmu zwanego *próbko-*
waniem (probierzem) Gibbsa (ang. Gibbs sampler, Glauber dynamics) (9.3), który jest
używany do symulacji rozkładu prawdopodobieństwa π na przestrzeni stanów D^V , gdzie D
jest zbiorem możliwych wartości każdego wierzchołka pochodzącego ze zbioru V .

Algorytm 9.3 (Probierz Gibbsa).

1. Wylosuj $v \in V$ (z równym prawdopodobieństwem)
2. Wylosuj $X_{n+1}(v)$ zgodnie z warunkowym rozkładem π względem wszystkich wartości $X_n(w)$ dla $w \neq v$
3. $\forall w \neq v: X_{n+1}(w) = X_n(w)$

Przykład 9.4 (Losowe q -kolorowanie).

Niech dany będzie graf $G = (V, E)$ oraz pewien zbiór $q \geq 2$ kolorów. Przez *pokoloro-*
wanie grafu będziemy rozumieć jedno konkretne przypisanie kolorów wierzchołkom (jedno
z $[q]^V$ możliwych). *Kolorowaniem grafu* nazwiemy takie jego pokolorowanie, w którym dwa
sąsiadujące ze sobą wierzchołki nie mają przypisanego tego samego koloru. Nas interesować
będą *q -kolorowania grafu*, czyli takie kolorowania, które używają nie więcej niż q kolorów.

Oznaczmy przez $Z_{G,q}$ liczbę wszystkich q -kolorowań grafu G . Wówczas dla dowolnego
pokolorowania $\xi \in [q]^V$ bierzemy miarę probabilistyczną

$$\rho_{G,q}(\xi) = \begin{cases} \frac{1}{Z_{G,q}} & \text{jeżeli } \xi \text{ jest } q\text{-kolorowaniem} \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Czyli losowym q -kolorowaniem jest q -kolorowanie wybrane z jednakowym prawdopodo-
bieństwem ze zbioru wszystkich q -kolorowań, czyli zgodnie z rozkładem jednostajnym $\rho_{G,q}$.

Algorytm 9.5 (Probierz Gibbsa dla losowego q -kolorowania).

1. Wylosuj $v \in V$ (z równym prawdopodobieństwem)
2. Wylosuj $X_{n+1}(v)$ zgodnie z warunkowym rozkładem $\rho_{G,q}$ względem wszystkich wartości $\{X_n(w) : w \neq v\}$
3. $\forall w \neq v: X_{n+1}(w) = X_n(w)$

Powstały w ten sposób łańcuch Markowa jest nieredukowalny i nieokresowy z rozkładem
stacjonarnym $\rho_{G,q}$, przy czym nie jest łatwym zadaniem pokazanie nieredukowalności
(zadanie domowe 8.2).

Inną wersją Algorytmu 9.3 dla losowego q -kolorowania jest tzw. wersja Metropolis:

Algorytm 9.6 (Wersja Metropolis dla losowego q -kolorowania).

1. Wylosuj $v \in V$ oraz $c \in [q]$ (oba z równym prawdopodobieństwem)
2. Jeżeli $\forall w \in N_G(v): X_n(w) \neq c$ (wierzchołek v można przekolorować na kolor c), to $X_{n+1}(v) = c$;
w przeciwnym przypadku $X_{n+1}(v) = X_n(v)$
3. $\forall w \neq v: X_{n+1}(w) = X_n(w)$

9.3. Przybliżone przeliczanie

Nawiązując do Przykładu 9.4 spróbujemy wejść w zagadnienie przeliczania interesujących nas obiektów za pomocą losowych algorytmów, które są szybkie i dają dobre przybliżenie z bardzo dużym prawdopodobieństwem.

Niech dany będzie graf $G = (V, E)$ oraz $q \geq 2$ kolorów – jak w Przykładzie 9.4. Naturalnym pytaniem, jakie się pojawia jest: *Ile jest różnych q -kolorowań danego grafu G ?*

Rozwiązanie naiwne sprowadza się do wygenerowania wszystkich q -kolorowań dla danego grafu G i ich zliczenia. Jest to jednak podejście niepraktyczne, zwłaszcza w przypadku „dużych” grafów, ze względu na bardzo dużą, wykładniczą liczbę kolorowań. Celem jest zatem znalezienie losowego algorytmu zdefiniowanego następująco.

Definicja 9.7. Algorytm losowy nazywamy *RPTAS* (z ang. Randomized Polynomial Time Approximation Scheme) jeśli spełnia następujące własności:

- (a) wielomianowy czas działania względem k (rozmiaru grafu)
- (b) dla każdego $\varepsilon > 0$ zwracany wynik z spełnia nierówność

$$(1 - \varepsilon) \cdot Z_{G,q} \leq z \leq (1 + \varepsilon) \cdot Z_{G,q}$$

- (c) prawdopodobieństwo sukcesu (czyli spełnienia poprzednich podpunktów) wynosi co najmniej $\frac{2}{3}$.

Definicja 9.8. Algorytmy, do których w podpunkcie (a) Definicji 9.7 wprowadzimy założenie, iż czas działania algorytmu powinien być wielomianowy względem $\frac{1}{\varepsilon}$, nazywamy *FPRAS* (z ang. Fully Polynomial Randomized Approximation Scheme).

Przykład 9.9 (Naiwne podejście do problemu q -kolorowania).

Niech dany będzie graf $G = (V, E)$, $|V| = k$ oraz $Z_{G,q}$ oznacza liczbę q -kolorowań w G . Przypuśćmy, że każdemu wierzchołkowi został przypisany kolor ze zbioru $[q]$ z jednakowym prawdopodobieństwem. Wówczas każda konfiguracja $\xi \in [q]^V$ powstaje z prawdopodobieństwem $\frac{1}{q^k}$. Zatem prawdopodobieństwo, że podczas tej procedury powstanie q -kolorowanie jest równe $\frac{Z_{G,q}}{q^k}$.

Rozważmy powtórzenie n razy powyższej procedury. W takim przypadku otrzymamy schemat Bernoulliego n prób z prawdopodobieństwem sukcesu $p = \frac{Z_{G,q}}{q^k}$. Niech Y_n będzie liczbą właściwych q -kolorowań otrzymanych w n próbach. Wówczas Y_n ma rozkład dwumianowy z $\mathbb{E}(Y_n) = n \cdot \frac{Z_{G,q}}{q^k}$, co implikuje, że $\mathbb{E}\left(\frac{q^k}{n} Y_n\right) = Z_{G,q}$, czyli $\frac{q^k}{n} Y_n$ jest nieobciążonym estymatorem dla $Z_{G,q}$. Zauważmy, że na podstawie nierówności Czebyszewa¹ mamy

$$\mathbb{P}\left(\left|\frac{q^k}{n} Y_n - Z_{G,q}\right| \geq \varepsilon \cdot Z_{G,q}\right) \leq \frac{q^{2k} \cdot \text{Var}(Y_n)}{n^2 \varepsilon^2 Z_{G,q}^2} \leq \frac{q^k}{n \varepsilon^2 Z_{G,q}}$$

gdzie $\text{Var}(Y_n) = n \frac{Z_{G,q}}{q^k} \left(1 - \frac{Z_{G,q}}{q^k}\right)$ jest wariancją zmiennej losowej Y_n , dla której zastosowano oszacowanie $1 - \frac{Z_{G,q}}{q^k} \leq 1$.

¹ $\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$

Aby były spełnione podpunkty (b) i (c) Definicji 9.7, (aproksymacja z błędem ε zachodziła z prawdopodobieństwem co najmniej $2/3$), trzeba założyć, że $n \geq 3 \frac{q^k}{\varepsilon^2 Z_{G,q}}$, co, na podstawie zadania domowego 8.3, jest co najmniej wykładniczą funkcją k (o podstawie $(q-1)/q$). Także nasz algorytm nie spełnia punktu (a) Definicji 9.7.

Zatem okazuje się, że zastosowanie zaproponowanej naiwnej metody nie jest dobrym rozwiązaniem problemu i trzeba zastosować bardziej wyrafinowany sposób, który zostanie przedstawiony na kolejnym wykładzie.

10. Wykład 10 i 11: Aproksymacja liczby kolorowań grafu

Wstęp

Podczas poprzedniego wykładu zajmowaliśmy się metodami Monte Carlo, które w sposób istotny korzystały z łańcuchów Markowa. Przyglądaliśmy się między innymi zagadnieniom przeliczania różnych obiektów za pomocą losowych algorytmów – szybkich i dających dobre przybliżenie z bardzo dużym prawdopodobieństwem. Najważniejszym pytaniem jakie sobie postawiliśmy było pytanie o liczbę różnych q -kolorowań danego grafu G . Rozpatrywany przez nas naiwny algorytm okazał się narzędziem, które niestety nie działa dostatecznie szybko z zadowalającą dokładnością. Podczas tego wykładu udowodnimy twierdzenie o istnieniu algorytmu klasy RPTAS (a nawet FPRAS) zliczającego wszystkie możliwe q -kolorowania dowolnego grafu i skonstruujemy taki algorytm.

Twierdzenie 10.1.

Dla wszystkich $\Delta > 2$ i $q > 4\Delta$ istnieje algorytm klasy RPTAS obliczający $Z_{G,q}$ dla wszystkich G takich, że $\Delta(G) \leq \Delta$.

10.1. Dowód istnienia algorytmu RPTAS dla obliczania liczby q -kolorowań grafu

Na początku zdefiniujemy, jak taki algorytm miałby wyglądać, a potem udowodnimy jego zgodność z Definicją RPTAS (9.7), a także, jak się okaże, zgodność z Definicją FPRAS (9.8).

Algorytm 10.2.

Dane wejściowe: $\varepsilon > 0$, $\Delta > 2$, $q > 4\Delta$, graf $G = (V, E)$ taki, że $\Delta(G) \leq \Delta$, $|V| = k$, $|E| = l$

Dane wyjściowe: $Y^* \in \mathbb{R}$: $\mathbb{P}(|Y^* - Z_{G,q}| \leq \varepsilon Z_{G,q}) \geq \frac{2}{3}$

1. Ponumeruj krawędzie grafu G : $E = (e_1, e_2, \dots, e_l)$. Niech $e_j = \{x_j, y_j\}$.
2. Skonstruuj rosnący ciąg grafów G_0, \dots, G_l takich, że dla $0 \leq j \leq l$:
 - $G_0 = (V, \emptyset)$
 - $G_j = (V, \{e_1, \dots, e_j\})$
 - $G_l = G$
3. Dla każdego $j = 1, 2, \dots, l$ wykonaj:
4. $Y_j := 0$
5. Dla każdego $i = 1, 2, \dots, m$ wykonaj:
6. Zapuć Probiez Gibbsa dla losowego q -kolorowania (Algorytm 9.4) na grafie G_{j-1} zaczynając z ustalonego stanu X_0 i zatrzymaj po n krokach
7. Jeśli $X_n(x_j) \neq X_n(y_j)$ to $Y_j := Y_j + \frac{1}{m}$

$$8. Y^* = q^k \prod_{j=1}^l Y_j$$

Żeby udowodnić nasze twierdzenie należy pokazać, że powyższy algorytm jest typu RPTAS, czyli, że dla każdego $\varepsilon > 0$ możemy znaleźć takie wartości n i m , że spełnione są następujące warunki zgodne z Definicją 9.7:

- (a) Czas działania względem rozmiaru grafu jest wielomianowy – wartość $O(lmn)$ jest asymptotycznie wielomianowa.
- (b) Zwracany wynik spełnia nierówność $|Y^* - Z_{G,q}| \leq \varepsilon Z_{G,q}$.
- (c) prawdopodobieństwo sukcesu (czyli spełnienia poprzednich podpunktów) wynosi co najmniej $\frac{2}{3}$.

Zajmijmy się podpunktem (b), czyli sprawdzeniem występującej tam nierówności. Przekształćmy ją do bardziej przyjaznej postaci:

$$1 - \varepsilon \leq \frac{Y^*}{Z_{G,q}} \leq 1 + \varepsilon$$

Aby to pokazać wprowadźmy pewne wygodne oznaczenia. Niech $Z_j = Z_{G_j,q}$ będzie liczbą q -kolorowań grafu G_j . Wówczas $Z_0 = q^k$, a szukaną przez nas liczbą będzie Z_l , którą możemy wyrazić w następujący sposób:

$$Z_l = \frac{Z_l}{Z_{l-1}} \frac{Z_{l-1}}{Z_{l-2}} \cdots \frac{Z_2}{Z_1} \frac{Z_1}{Z_0} Z_0$$

Wprowadźmy teraz następujący lemat:

Lemat 10.3:

Niech $0 \leq \varepsilon \leq 1$ będzie dowolne i niech $a_1, \dots, a_l, b_1, \dots, b_l$ będą liczbami rzeczywistymi większymi od zera takimi, że:

$$1 - \frac{\varepsilon}{2l} \leq \frac{a_j}{b_j} \leq 1 + \frac{\varepsilon}{2l}$$

Wówczas:

$$1 - \varepsilon \leq \prod_{i=1}^l \frac{a_i}{b_i} \leq 1 + \varepsilon$$

Dowód tego lematu jest treścią zadania domowego 8.4.

Z Lematu 10.3 wynika, że aby wykazać interesującą nas nierówność wystarczy udowodnić dla każdego $j = 1, 2, \dots, l$ nierówność:

$$1 - \frac{\varepsilon}{2l} \leq \frac{Y_j}{Z_{j-1}} \leq 1 + \frac{\varepsilon}{2l}$$

Wprowadźmy kolejny lemat pomocniczy:

Lemat 10.4:

Niech $\Delta \geq 2$ i niech $q \geq \Delta + 2$. Wtedy dla dowolnego grafu G takiego, że $\Delta(G) \leq \Delta$ (gdzie $\Delta(G)$ oznacza największy stopień wierzchołka w grafie) i dowolnego q -kolorowania X mamy:

$$\rho_{G,p}(X(x) \neq X(y)) \geq \frac{1}{2}$$

gdzie x i y są dowolnymi, różnymi wierzchołkami badanego grafu, a $\rho_{G,p}$ jest rozkładem jednostajnym na q -kolorowaniach.

Dowód. Jeśli x i y są połączone krawędzią to nie ma czego dowodzić, ponieważ zawsze będą one różnych kolorów (z prawdopodobieństwem $1 > \frac{1}{2}$). Rozważmy więc przypadek gdy x i y nie są połączone. Zgodnie ze wzorem na prawdopodobieństwo całkowite mamy:

$$\mathbb{P}(X(x) \neq X(y)) = \sum_{\chi(x) \in K} \mathbb{P}(X(x) \neq X(y) | \chi(x)) \mathbb{P}(\chi(x))$$

gdzie K jest zbiorem wszystkich możliwych q -kolorowań grafu G bez wierzchołka x , a $\chi(x)$ jest jednym z takich kolorowań. Oznaczmy przez c liczbę kolorów jaką może przybrać x tak aby dla wszystkich sąsiadów y tego wierzchołka $X(x) \neq X(y)$ przy ustalonych kolorach wszystkich wierzchołków poza x . Oczywiście $c \geq q - \Delta$. Możemy więc napisać:

$$\sum_{\chi(x) \in K} \mathbb{P}(X(x) \neq X(y) | \chi(x)) \mathbb{P}(\chi(x)) \geq \frac{c-1}{c} \sum_{\chi(x) \in K} \mathbb{P}(\chi(x)) \stackrel{\times}{=} \frac{c-1}{c} \geq \frac{q-\Delta-1}{q-\Delta}$$

gdzie równość \times wynika z tego, że $\sum_{\chi(x) \in K} \mathbb{P}(\chi(x))$ jest całą przestrzenią zdarzeń, więc sumuje się do jedynki. Wystarczy teraz przypomnieć sobie, że $q \geq \Delta + 2$ i składając wszystkie poprzednie nierówności mamy, że:

$$P(X(x) \neq X(y)) \geq \frac{q-\Delta-1}{q-\Delta} \geq \frac{1}{2}$$

□

Graf, który rozpatrujemy spełnia oczywiście założenia Lematu 10.4. Łatwo widać, że nierówność, którą próbujemy udowodnić, dzięki Lematowi 10.4 jest implikowana przez następującą:

$$\frac{\varepsilon}{4l} \leq Y_j - \frac{Z_j}{Z_{j-1}} \leq \frac{\varepsilon}{4l}$$

Przyjrzyjmy się zatem bliżej wyrażeniu $\frac{Z_j}{Z_{j-1}}$. Mianownik równy jest liczbie kolorowań grafu G_{j-1} , a licznik jest liczbą kolorowań grafu G_{j-1} , w których wierzchołki x_i i y_i mają różny kolor. Przyjmijmy zatem, że:

$$\frac{Z_i}{Z_{i-1}} = \rho_{G_{j-1},q}(\text{kolor}(x_i) \neq \text{kolor}(y_i))$$

gdzie $\rho_{G_{j-1},q}$ oznacza rozkład jednostajny na q -kolorowaniach. Zatem interesującą nas nierówność możemy zapisać następująco:

$$\begin{aligned} \left| Y_j - \frac{Z_j}{Z_{j-1}} \right| &= |Y_j - \rho(X(x_j) \neq X(y_j))| \stackrel{\text{nierówność trójkąta}}{\leq} \\ &\leq |Y_j - \mu^{(n)}(X(x_j) \neq X(y_j))| + |\mu^{(n)}(X(x_j) \neq X(y_j)) - \rho(X(x_j) \neq X(y_j))| \leq \frac{\varepsilon}{4l} \end{aligned}$$

Aby więc wykazać prawdziwość punktu (b) Definicji 9.7 wystarczy pokazać następujące nierówności:

$$|\mu^{(n)}(X(x_j) \neq X(y_j)) - \rho(X(x_j) \neq X(y_j))| \stackrel{*}{\leq} \frac{\varepsilon}{8l}$$

$$|Y_j - \mu^{(n)}(X(x_j) \neq X(y_j))| \stackrel{**}{\leq} \frac{\varepsilon}{8l}$$

Rozbiliśmy więc w ten sposób błąd popełniany przez nasz algorytm na jego dwie „składowe”. Nierówność $**$ oznacza błąd wynikły podczas estymacji Y_j przez kolejne wyniki próbkowania probierzem Gibbsa, natomiast nierówność $*$ oznacza błąd, który wynika bezpośrednio z probierza Gibbsa (tego, jak ona przybliża). (Rozwikłując nazwę MCMC, to właśnie $*$ jest częścią Markov Chain (łańcuchy Markowa), a $**$ jest częścią Monte Carlo.)

Wspomnijmy też o tym, że nierówności, które próbujemy tutaj udowodnić nie muszą być prawdziwe zawsze, a jedynie zachodzić dla pewnych m i n z prawdopodobieństwem co najmniej $\frac{2}{3}$ (mówi o tym punkt (c) naszego celu). To, co tak naprawdę trzeba dowiedzieć, to:

$$\mathbb{P}\left(|\mu^{(n)}(X(x_j) \neq X(y_j)) - \rho(X(x_j) \neq X(y_j))| \leq \frac{\varepsilon}{8l}\right) \stackrel{*}{\geq} \frac{2}{3}$$

$$\mathbb{P}\left(|Y_j - \mu^{(n)}(X(x_j) \neq X(y_j))| \leq \frac{\varepsilon}{8l}\right) \stackrel{**}{\geq} \frac{2}{3}$$

Naszym celem będzie teraz znalezienie takiego m , dla którego obie te nierówności zachodzą. (Przypomnijmy, że m mówi nam, ile razy należy wykonać probierz Gibbsa w Algorytmie 10.2, aby otrzymać dobry wynik.)

Nierówność $**$ udowodnimy przy pomocy następującego lematu:

Lemat 10.5:

Jeżeli zmienna losowa X ma rozkład $Bin(n, p)$ i $a > 0$ jest dowolne to:

$$\mathbb{P}(|X - np| \geq a) \leq \frac{n}{4a^2}$$

Dowód. Natychmiastowy wniosek z nierówności Czebyszewa. □

Przekształćmy nierówność $**$ w taki sposób, aby rozpatrywać w nim prawdopodobieństwo zdarzenia odwrotnego do danego. Wówczas mamy równoważną postać:

$$\mathbb{P}\left(|Y_j - \mu^{(n)}(X(x_j) \neq X(y_j))| > \frac{\varepsilon}{8l}\right) \stackrel{**}{<} \frac{1}{3}$$

Przypomnijmy, że nierówność ta mówi cały czas o prawdopodobieństwie, że probierz Gibbsa w Algorytmie 10.2 da nam wynik prawidłowy. Wykonujemy go m razy z prawdopodobieństwem sukcesu $p = \mu^{(n)}(X(x_j) \neq X(y_j))$, czyli jest to rozkład dwumianowy $Bin(m, p)$. Możemy zatem użyć Lematu 10.5. Pomnóżmy nierówność pod prawdopodobieństwem przez szukane m i przekształćmy lewą stronę nierówności korzystając z Lematu 10.5:

$$\mathbb{P}\left(|mY_j - m\mu^{(n)}(X(x_j) \neq X(y_j))| > \frac{m}{\varepsilon}8l\right) \stackrel{10.5}{\leq} \frac{64ml^2}{4m^2\varepsilon^2} = \frac{16l^2}{m\varepsilon^2}$$

Zatem nierówność $**$ będzie prawdziwa, o ile:

$$\frac{16l^2}{m\varepsilon^2} < \frac{1}{3l}$$

$$m > \frac{48l^3}{\varepsilon^2}$$

l pojawiło się tu dlatego, że musimy jeszcze uwzględnić liczbę m krotnego zapuszczania probierza Gibbsa.

Zauważmy jeszcze, że w naszym grafie zachodzi nierówność $l \leq \frac{\Delta k}{2}$, zatem za m , dla którego nierówność ** będzie prawdziwa, wystarczy wziąć:

$$m = \left\lceil \frac{6\Delta^3 k^3}{\varepsilon^2} \right\rceil$$

Wiemy już zatem ile razy musimy uruchamiać probierz Gibbsa w Algorytmie 10.2. Wyliczmy teraz jak długo każdy probierz Gibbsa musi działać, aby wynik zbiegł z dużym prawdopodobieństwem ($> \frac{2}{3}$) do wyniku przez nas oczekiwanego, czyli rozkładu stacjonarnego. O tym właśnie mówi nierówność *.

Wprowadźmy oznaczenie:

$$\tau(\varepsilon) := \max_{X_0} \min\{n : d_{TV}(\mu_s^{(n)}, \pi) \leq \varepsilon\}$$

gdzie maksimum bierzemy po wszystkich stanach początkowych probierza Gibbsa, a π oznacza rozkład stacjonarny. Wartość $\tau(\varepsilon)$ mówi o minimalnej liczbie iteracji wymaganej do osiągnięcia dokładności ε . Oznacza to, że jeśli weźmiemy

$$n = \tau\left(\frac{\varepsilon}{8l}\right)$$

to zapewnimy sobie odpowiednią zbieżność naszego probierza Gibbsa w Algorytmie 10.2.

Udowodnijmy zatem następujące twierdzenie:

Twierdzenie 10.6.

Dla dowolnego Δ dla każdego grafu G takiego, że $\Delta(G) \leq \Delta$ i $q \geq 4\Delta + 1$ łańcuch Markowa probierza Gibbsa dla q -kolorowania grafu G (9.4) spełnia nierówność:

$$\tau(\varepsilon) \leq \frac{kq}{q - 4\Delta} \ln\left(\frac{k}{\varepsilon}\right)$$

Dowód. W dowodzie tego twierdzenia skorzystamy z nierówności $d_{TV}(X, Y) \leq \mathbb{P}(X \neq Y)$ i metody couplingu. Przypomnijmy, że couplingiem dla łańcucha Markowa M nazywamy łańcuch Markowa $Z = (X, Y)$ zdefiniowany na S^2 taki, że:

- (i) $X_0 = \xi_0$, a Y_0 jest wylosowany zgodnie z rozkładem π
- (ii) $\mathbb{P}(X_{n+1} = s' | Z_n = (s, t)) = \mathbb{P}(M_{n+1} = s' | M_n = s)$
- (iii) $\mathbb{P}(Y_{n+1} = t' | Z_n = (s, t)) = \mathbb{P}(M_{n+1} = t' | M_n = t)$

Warunki (ii) i (iii) oznaczają, że każda ze współrzędnych zmiennej Z ma taki sam rozkład jak wyjściowy łańcuch Markowa M . W naszym wypadku coupling zdefiniujemy następująco: przejście ze stanu (X_n, Y_n) dokonuje się poprzez wylosowanie z rozkładem jednostajnym wierzchołka v i koloru c jednocześnie dla obu łańcuchów. Niech teraz $D_n = \{v : X_n(v) \neq$

$Y_n(v)$, a $A_n = V \setminus D_n$. Chcemy by $D_n = \emptyset$, bo wtedy łańcuch osiągnie stan stacjonarny i już w nim pozostanie. Mamy z nierówności Markowa:

$$\mathbb{P}(X_n \neq Y_n) = \mathbb{P}(|D_n| \geq 1) \leq \mathbb{E}(|D_n|)$$

Wykorzystamy dalej równość $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$. Oznaczmy przez $d_n = |D_n|$. Mamy:

$$\begin{aligned} & \mathbb{E}(d_{n+1}|d_n = d) = \\ &= \mathbb{P}(d_{n+1} = d + 1|d_n = d)(d + 1) + \mathbb{P}(d_{n+1} = d|d_n = d)d + \mathbb{P}(d_{n+1} = d - 1|d_n = d)(d - 1) = \\ &= d + \mathbb{P}(d_{n+1} = d + 1|d_n = d) - \mathbb{P}(d_{n+1} = d - 1|d_n = d) \end{aligned}$$

Stopień v wynosi co najwyżej Δ , więc pozostaje co najmniej $q - 2\Delta$ innych kolorów na sąsiednich wierzchołkach w obu grafach. Stąd:

$$\begin{aligned} & \mathbb{P}(d_{n+1} = d - 1|d_n = d)(d - 1) \geq \frac{d(q - 2\Delta)}{kq} \\ & \mathbb{P}(d_{n+1} = d + 1|d_n = d)(d + 1) \leq \frac{\sum_{v \in D_n} 2 \deg(v, A_n)}{kq} = \frac{\sum_{v \in A_n} 2 \deg(v, D_n)}{kq} \leq \frac{2d\Delta}{kq} \end{aligned}$$

Na podstawie powyższych:

$$\mathbb{E}(d_{n+1}|d_n = d) \leq d + \frac{2d\Delta}{kq} - \frac{d(q - 2\Delta)}{kq} = d \left(1 - \frac{q - 4\Delta}{kq}\right)$$

Korzystając z równości dla wartości oczekiwanej otrzymujemy:

$$\mathbb{E}(d_{n+1}) = \mathbb{E}(\mathbb{E}(d_{n+1}|d_n = d)) \leq \mathbb{P}(d_n) \left(1 - \frac{q - 4\Delta}{kq}\right)$$

Z indukcji dla powyższej zależności dostajemy:

$$\mathbb{P}(d_n) \leq \left(1 - \frac{q - 4\Delta}{kq}\right)^n \mathbb{P}(d_0)$$

A ponieważ $\mathbb{P}(d_0) \leq k$, zaś wyrażenie w nawiasie można ograniczyć korzystając z nierówności $e^{-x} \geq 1 - x$ dla $x \in [0, 1]$ szacujemy ostatecznie:

$$\mathbb{P}(d_n) \leq k e^{-\frac{q-4\Delta}{kq}n} \leq \varepsilon$$

Czyli d_{TV} wynosi co najwyżej ε po $n = \lceil \frac{kq}{q-4\Delta} \ln \left(\frac{k}{\varepsilon}\right) \rceil$ krokach co kończy dowód. \square

Zatem z Twierdzenia 10.6 wynika, że za n w Algorytmie 10.2 trzeba wziąć wartość

$$n = \left\lceil \tau \left(\frac{\varepsilon}{8l}\right) \right\rceil = \left\lceil \frac{kq}{q - 4\Delta} \ln \left(\frac{k}{8l}\right) \right\rceil$$

Przypomnimy, że w naszym grafie zachodzi nierówność $l \leq \frac{\Delta k}{2}$, zatem za n , dla którego nierówność * będzie prawdziwa, wystarczy wziąć:

$$n = \left\lceil \frac{kq}{q - 4\Delta} \ln \left(\frac{1}{4\Delta}\right) \right\rceil$$

Pokazaliśmy zatem część (b) z prawdopodobieństwem określonym w (c) definicji RPTAS (9.7). Została nam tylko część (a), czyli złożoność algorytmu. Policzmy ją. Wykonujemy l razy m próbkowań Gibbsa, a w każdym próbkowaniu łańcuch Markowa wykonuje n przejść. Zatem czas wykonywania algorytmu wynosi $l \cdot m \cdot n$. Podstawmy więc obliczone przez nas wartości i policzmy złożoność obliczeniową:

$$l \cdot m \cdot n = \frac{k\Delta}{2} \cdot \frac{6\Delta^3 k^3}{\varepsilon^2} \cdot \frac{kq}{q-4\Delta} \ln\left(\frac{1}{4\Delta}\right) = O\left(k \cdot \frac{k^3}{\varepsilon^2} \cdot k\right) = O\left(\frac{k^5}{\varepsilon^2}\right)$$

Dzięki temu ostatecznie udowodniliśmy, że Algorytm 10.2 jest klasy RPTAS. Co więcej, takie oszacowanie złożoności oznacza, że tak jak zapowiedzieliśmy nasz algorytm jest również klasy FPRAS, gdyż jest wielomianowy ze względu na k i odwrotność ε (z Definicji 9.8).

10.2. Poprawienie oszacowania z Twierdzenia 10.6

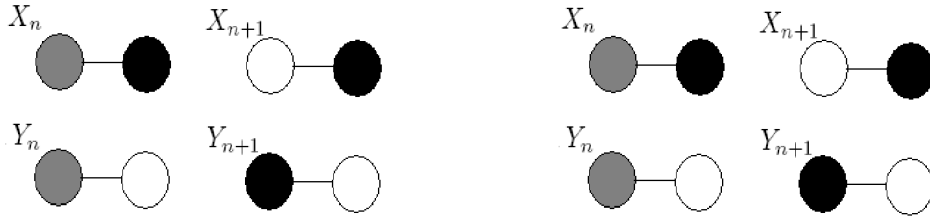
W twierdzeniu 10.6 można osłabić nieco założenia oraz poprawić oszacowanie. Mówi o tym następujące twierdzenie:

Twierdzenie 10.7.

Dla dowolnego Δ dla każdego grafu G takiego, że $\Delta(G) \leq \Delta$ i $q \geq 2\Delta + 1$ łańcuch Markowa probierza Gibbsa dla q -kolorowania grafu G (9.4) spełnia nierówność:

$$\tau(\varepsilon) \leq \frac{kq}{q-2\Delta} \ln\left(\frac{k}{\varepsilon}\right)$$

Dowód. Wprowadźmy oznaczenia dla D_n , d_n i A_n , takie same jak w dowodzie Twierdzenia 10.6. Dla wierzchołków v należących do A_n przez $d'(v)$ oznaczmy liczbę wierzchołków przyległych do v i należących do D_n , odwrotnie dla $w \in D_n$ przez $d'(w)$ rozumiemy liczbę wierzchołków przyległych należących do A_n .



Rysunek 10.1. Po lewej stronie przedstawiony został oryginalny coupling, po prawej zaś poprawiony. W wersji oryginalnej couplingu szary wierzchołek ma ten sam kolor w obu łańcuchach, a jego sąsiad jest w innym kolorze (odpowiednio czarny i biały). Przy próbie przekolorowania szarego wierzchołka na czarny będzie to możliwe w jednym łańcuchu lecz niemożliwe w drugim, zwiększając d_n . Podobna sytuacja nastąpi przy próbie kolorowania na biało, po raz kolejny zwiększając d_n . W wersji poprawionej couplingu, jeśli szary wierzchołek ma zostać przekolorowany na biało w X_n to zostaje przemalowany na czarno w Y_n i na odwrót, tylko raz zwiększając d_t .

Rozważmy następujący coupling: jeżeli wierzchołek $v \in D_n$ został wylosowany do przekolorowania to stosujemy ten sam kolor dla dwóch łańcuchów (tzn. postępujemy jak poprzed-

nio). Kolor v nie zmieni się, jeśli wybrany kolor będzie różny od kolorów w obu łańcuchach. Mamy $c - 2\Delta + d'(v)$ takich kolorów (ograniczamy tu bardziej!). Zatem:

$$\mathbb{P}(d_{n+1} = d - 1 | d_n = d) = \sum_{v \in D_n} \frac{q - 2\Delta + d'(v)}{kq}$$

Jeżeli krawędź, która ma być pokolorowana należy do A_n modyfikujemy coupling. Zauważmy, że gdyby v miało kolor zielony, a sąsiad w byłby czerwony w jednym łańcuchu, zaś niebieski w drugim łańcuchu i ponadto żaden inny sąsiad nie miałby tych kolorów, wówczas próbując przekolorować v na któryś z nich spowodujemy, że w jednym łańcuchu kolor v się zmieni, a drugim nie. Są więc dwie możliwości wyboru koloru, co zwiększa d_n . W tej szczególnej sytuacji próbujemy zamalowywać v kolorami pochodzącymi od przeciwnego łańcucha. Wówczas w obu kolor się zmieni albo w obu nie zmieni. W ten sposób mniej zwiększamy d_n . Ogólniej, niech S_1 będzie zbiorem kolorów sąsiadów w pierwszym łańcuchu, a zarazem nie będącymi tymi kolorami w drugim. Natomiast S_2 oznacza zbiór kolorów sąsiadów w drugim, ale nie w pierwszym łańcuchu. Sparujmy kolory po jednym z każdego zbioru $c_1 \in S_1$ i $c_2 \in S_2$ na ile to możliwe i tak jak w przypadku dla pojedynczej pary wymieniamy kolory w obrębie tych par przy każdym umożliwiającym to losowaniu. W wyniku tego całkowita liczba możliwości w kolorowaniu wzrośnie co najwyżej o $\max(S_1(v), S_2(v))$. Mamy więc:

$$\mathbb{P}(d_{n+1} = d + 1 | d_n = d) = \sum_{v \in A_n} \frac{d'(v)}{kq}$$

I stąd postępując jak w dowodzie Twierdzenia 10.6 otrzymujemy:

$$\mathbb{E}(d_{n+1} | d_n = d) \leq d \left(1 - \frac{q - 2\Delta}{kq} \right)$$

$$\mathbb{E}(d_n) \leq ke^{-\frac{q-2\Delta}{kq}n} \leq \varepsilon$$

Czyli d_{TV} wynosi co najwyżej ε po $n = \left\lceil \frac{kq}{q-2\Delta} \ln\left(\frac{k}{\varepsilon}\right) \right\rceil$ krokach co kończy dowód. \square

11. Wykład 12: Martynały: definicja, twierdzenia o zbieżności

Wstęp

Do tej pory zajmowaliśmy się ciągami zmiennych losowych (X_n) o pewnej strukturze zależności. Ciąg zmiennych losowych niezależnych stanowił przypadek szczególny. Badaliśmy pewne własności łańcuchów Markowa, tj. takich ciągów zmiennych losowych, dla których kolejny wyraz zależał wyłącznie od poprzedniego (tzw. własność zaniku pamięci). Niniejszy wykład ma na celu przedstawienie nowej struktury zależności dla ciągów zmiennych losowych, jaką określają martynały.

Pierwsza część wykładu poświęcona jest przypomnieniu wiadomości na temat warunkowej wartości oczekiwanej. Zawiera ona definicję oraz kilka podstawowych twierdzeń. W kolejnej części wprowadzone zostaje pojęcie martynału. Przedstawiono tu również kilka przykładów. Ostatnia, trzecia część wykładu, dotyczy twierdzeń o zbieżności martynałów.

11.1. Warunkowa wartość oczekiwana

11.1.1. Przypadek jednowymiarowy

Wszelkie rozważania dotyczące warunkowej wartości oczekiwanej ograniczone zostaną do zmiennych losowych dyskretnych.

Definicja 11.1. *Warunkową wartością oczekiwaną zmiennej losowej Y pod warunkiem $X = x$ nazywamy liczbę $\psi(x)$ określoną następująco:*

$$\psi(x) = E(Y|X = x) = \sum_y y \mathbb{P}(Y = y|X = x)$$

Zauważmy, że $\psi(X)$ jest pewną funkcją zmiennej losowej X . Zatem $\psi(X)$ jest również zmienną losową. Oznaczamy ją przez $\psi(X) = \mathbb{E}(Y|X)$. Wartość oczekiwana tak określonej zmiennej losowej jest równa $\mathbb{E}(Y)$, o czym mówi następujące twierdzenie:

Twierdzenie 11.2.

$$\mathbb{E}(\psi(X)) = \mathbb{E}(Y)$$

Dowód.

$$\begin{aligned}\mathbb{E}(\psi(X)) &= \sum_x \psi(x) \mathbb{P}_X(x) = \sum_x \sum_y y \mathbb{P}_{Y|X}(y|x) \mathbb{P}_X(x) = \sum_x \sum_y y \mathbb{P}_{X,Y}(x,y) = \\ &= \sum_y y \sum_x \mathbb{P}_{X,Y}(x,y) = \sum_y y \mathbb{P}_Y(y) = \mathbb{E}(Y).\end{aligned}$$

□

Powyższe twierdzenie można uogólnić:

Twierdzenie 11.3.

$$\mathbb{E}(\psi(X)g(X)) = \mathbb{E}(\psi(g(X)))$$

Dowód. Dowód tego faktu przebiega analogicznie jak Twierdzenia 11.2.

□

11.1.2. Przypadek wielowymiarowy

Zajmiemy się teraz przypadkiem, gdy zmienna losowa \mathbf{X} jest postaci $\mathbf{X} = (X_1, X_2, \dots, X_n)$.

Lemat 11.4:

Zachodzą następujące własności:

- (a) $\mathbb{E}(Y_1 + Y_2 | \mathbf{X}) = \mathbb{E}(Y_1 | \mathbf{X}) + \mathbb{E}(Y_2 | \mathbf{X})$
- (b) $\mathbb{E}(Yg(\mathbf{X}) | \mathbf{X}) = g(\mathbf{X})\mathbb{E}(Y | \mathbf{X})$
- (c) Jeśli h jest funkcją różnowartościową, to $\mathbb{E}(Y|h(\mathbf{X})) = \mathbb{E}(Y | \mathbf{X})$

Lemat 11.5 (Własność wieżowa):

$$\mathbb{E}(\mathbb{E}(Y | \mathbf{X}_1, \mathbf{X}_2) | \mathbf{X}_1) = \mathbb{E}(Y | \mathbf{X}_1).$$

Zdefiniujemy teraz warunkową wartość oczekiwaną zmiennej losowej Y względem zdarzenia A .

Definicja 11.6. Niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie przestrzenią probabilistyczną, na której określone są zmienne losowe X oraz Y . *Warunkową wartością oczekiwaną zmiennej losowej Y względem zdarzenia $A \in \mathcal{F}$ nazywamy liczbę*

$$\mathbb{E}(Y|A) = \sum_y y \mathbb{P}(Y = y|A),$$

gdzie $\mathbb{P}(Y = y|A) = \mathbb{P}(Y = y | \{X(\omega) : \omega \in A\})$.

Mamy

$$\mathbb{E}(Y|I_A) = \begin{cases} \mathbb{E}(Y|A) & \text{gdy } \omega \in A \\ \mathbb{E}(Y|A^c) & \text{gdy } \omega \notin A. \end{cases}$$

Zatem $\mathbb{E}(Y|I_A)$ jest zmienną losową dwupunktową. Ponadto zachodzi zależność $\mathbb{E}(I_B|A) = \mathbb{P}(B|A)$

Lemat 11.7:

Niech A będzie zdarzeniem oraz niech $A = \bigcup_{i=1}^n B_i$, gdzie $B_i \in \mathcal{F}$ są zdarzeniami parami rozłącznymi. Wówczas

$$\mathbb{E}(Y|A) \mathbb{P}(A) = \sum_{i=1}^n \mathbb{E}(Y|B_i) \mathbb{P}(B_i).$$

W szczególności, dla $A = \Omega$ z Lematu 11.7 wynika uogólnienie wzoru na prawdopodobieństwo całkowite:

$$\mathbb{E}(Y) = \sum_{i=1}^n \mathbb{E}(Y|B_i) \mathbb{P}(B_i).$$

Podstawiając w powyższym wzorze $Y = I_C$ otrzymujemy znany wzór:

$$\mathbb{P}(C) = \mathbb{E}(I_C) = \sum_{i=1}^n \mathbb{E}(I_C|B_i) \mathbb{P}(B_i) = \sum_{i=1}^n \mathbb{P}(C|B_i) \mathbb{P}(B_i).$$

Dowód Lematu 11.7. Z definicji wartości oczekiwanej wiemy, że:

$$\begin{aligned} \mathbb{E}(YI_C) &= \sum_y (y \mathbb{P}(Y = y, C)) = \sum_y (y \mathbb{P}(Y = y|C) \mathbb{P}(C)) = \mathbb{P}(C) \sum_y (y \mathbb{P}(Y = y|C)) = \\ &= \mathbb{P}(C) \mathbb{E}(Y|C) \end{aligned}$$

gdzie ostatnia równość wynika z definicji $\mathbb{E}(Y|C)$. Mamy zatem

$$\mathbb{E}(Y|A) \mathbb{P}(A) = \mathbb{E}(YI_A) = \mathbb{E}\left(Y \sum_{i=1}^n I_{B_i}\right) = \sum_{i=1}^n \mathbb{E}(YI_{B_i}) = \sum_{i=1}^n \mathbb{E}(Y|B_i) \mathbb{P}(B_i).$$

□

Definicja 11.8. Niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie przestrzenią probabilistyczną, $\mathcal{G} \subseteq \mathcal{F}$ pod σ -ciałem σ -ciała \mathcal{F} . Niech ponadto $Y : \Omega \rightarrow \mathbb{R}$ będzie zmienną losową taką, że $\mathbb{E}(Y^2) < \infty$. \mathcal{G} -mierzalną zmienną losową Z nazywamy *warunkową wartością oczekiwaną* względem \mathcal{G} , jeżeli

$$\mathbb{E}((Y - Z)I_G) = 0$$

dla każdego $G \in \mathcal{G}$. Oznaczamy ją $Z = \mathbb{E}(Y|\mathcal{G})$.

11.2. Martyngały, definicja i przykłady

W tej części wykładu wprowadzona zostanie definicja martyngału oraz podanych będzie kilka przykładów.

Definicja 11.9. Ciąg (S_n) zmiennych losowych (skończony lub nie) jest *martyngałem* względem ciągu (X_n) , jeżeli

- (a) $\mathbb{E}(|S_n|) < \infty$,
- (b) $\mathbb{E}(S_{n+1}|X_1, X_2, \dots, X_n) = S_n$.

Martynał określa zatem tzw. grę sprawiedliwą w takim sensie, że średnia wygrana w chwili $n + 1$, gdy znany jest przebieg gry do chwili n , jest równa S_n , czyli łącznej wygranej w chwili n .

Często definiuje się $S_n = X_n$ lub określa się S_n jako pewną funkcję X_n , tj. $S_n = \phi(X_n)$ (z definicji natomiast mamy $S_n = \phi(X_1, X_2, \dots, X_n)$).

Przykład 11.10. Niech X_1, X_2, \dots będą niezależnymi zmiennymi losowymi takimi, że $\mathbb{E}(X_i) = 0$ oraz $\mathbb{E}(|X_i|) < \infty$. Zdefiniujmy

$$S_n = X_1 + X_2 + \dots + X_n.$$

Zachodzi zależność

$$\begin{aligned} \mathbb{E}(S_{n+1}|X_1, X_2, \dots, X_n) &= \mathbb{E}(S_n|X_1, X_2, \dots, X_n) + \mathbb{E}(X_{n+1}|X_1, X_2, \dots, X_n) = \\ &= S_n + \mathbb{E}(X_{n+1}) = S_n \end{aligned}$$

Zatem (S_n) jest zatem martynałem względem ciągu (X_n) .

Przykład 11.11. Do warunków z poprzedniego przykładu dodajmy $\text{Var}(X_i) < \infty$. Niech $T_n = S_n^2$. Mamy wtedy

$$\mathbb{E}(T_{n+1}|X_1, X_2, \dots, X_n) = T_n + 2\mathbb{E}(X_{n+1})\mathbb{E}(S_n|X_1, X_2, \dots, X_n) + \mathbb{E}(X_{n+1}^2) \geq T_n.$$

Zatem (T_n) nie jest zatem martynałem względem ciągu (X_n) . Gdy spełniony jest taki rodzaj zależności, mówimy, że (T_n) jest *supermartynałem* względem ciągu (X_n) .

Przykład 11.12. Rozważmy prosty spacer losowy (zobacz Definicję 1.1), dla którego $\mathbb{P}(X_n = 1) = p$ i $\mathbb{P}(X_n = -1) = 1 - p = q$. Niech

$$S_0 = 0 \text{ oraz } S_n = \sum_{i=1}^n X_i.$$

Sprawdźmy, czy (S_n) jest martynałem względem (X_n) . Mamy $|S_n| \leq n$. Stąd $\mathbb{E}(|S_n|) < \infty$. Ponadto

$$\mathbb{E}(S_{n+1}|X_1, X_2, \dots, X_n) = S_n + p - q.$$

Zatem (S_n) nie jest martynałem. Zdefiniujmy ciąg (Y_n) następująco:

$$Y_n = S_n - \mathbb{E}(S_n) = S_n - n(p - q).$$

Wtedy $\mathbb{E}(|Y_n|) < \infty$ oraz

$$\mathbb{E}(Y_{n+1}|X_1, X_2, \dots, X_n) = S_n + p - q - (n + 1)(p - q) = S_n - n(p - q) = Y_n,$$

czyli (Y_n) jest martynałem względem ciągu (X_n) .

Przykład 11.13. Rozważmy pewną grę. Niech S_0 oznacza kapitał początkowy, S_n kapitał po n grach. Gra jest sprawiedliwa w potocznym sensie gdy

$$\mathbb{E}(S_{n+1}|S_0, S_1, \dots, S_n) = S_n,$$

czyli ciąg (S_n) jest martyngałem względem samego siebie. Załóżmy, że gracz stosuje strategię podwajania stawki po każdej przegranej i gra do czasu uzyskania pierwszego sukcesu. Z prawdopodobieństwem równym 1 strategia ta przynosi sukces, tzn. gracz zyskuje dokładnie 1. Zastanowimy się teraz jaki powinien być kapitał początkowy gracza, aby odniósł on sukces z prawdopodobieństwem 1. Niech L będzie zmienną losową określającą łączną przegraną do momentu pierwszej wygranej, N natomiast niech oznacza liczbę odbytych gier. N jest zmienną losową o rozkładzie geometrycznym z parametrem $1/2$ (Wtedy $\mathbb{E}(N) = 2$.) Mamy:

$$\mathbb{E}(L) = \mathbb{E}(\mathbb{E}(L|N)) = \sum_{n=2}^{\infty} \left(\frac{1}{2}\right)^{n-1} \cdot \frac{1}{2}(1 + 2 + \dots + 2^{n-2}) = \infty.$$

Zatem, aby wygrać z prawdopodobieństwem równym jeden, należy przyjść z nieskończenie wielkim kapitałem początkowym. Strategia podwajania stawki określa historycznie pierwszy martyngał.

Przykład 11.14. Określimy teraz dwa martyngały na bazie procesu gałęzkiego (zobacz Rozdział 4.2), gdzie $Z_0 = 1$ oraz Z_n dla $n \geq 1$ są liczebnościami kolejnych pokoleń.

(a) Mamy $\mathbb{E}(Z_{n+1}|Z_n = z_n) = z_n \cdot \mu$, gdzie $\mu = \mathbb{E}(Z_0)$. Stąd

$$\mathbb{E}(Z_{n+1}|Z_1, Z_2, \dots, Z_n) = \mathbb{E}(Z_{n+1}|Z_n) = \mu Z_n.$$

Ponadto $\mathbb{P}(Z_n) = \mu^n$. Zdefiniujmy

$$W_n = \frac{Z_n}{\mu^n}.$$

Wówczas

$$\mathbb{E}(W_{n+1}|Z_1, Z_2, \dots, Z_n) = \frac{\mu Z_n}{\mu^n} = W_n$$

i (W_n) jest martyngałem względem (Z_n) .

(b) Niech η będzie prawdopodobieństwem wyginięcia procesu gałęzkiego. Wtedy ciąg (V_n) określony wzorem $V_n = \eta^{Z_n}$ jest martyngałem względem (Z_n) .

11.3. Twierdzenia o zbieżności martyngałów

Na początku tej części wykładu przypomnimy definicje kilku typów zbieżności zmiennych losowych.

Definicja 11.15. Mówimy, że X_n dąży do X z prawdopodobieństwem jeden, jeżeli

$$\mathbb{P}\left(\left\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Stosujemy wtedy oznaczenie $X_n \xrightarrow{a.s.} X$.

Definicja 11.16. X_n zbiega do X według r -tego momentu, gdy $\mathbb{E}(X_n)^r < \infty$ oraz

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^r) = 0.$$

Stosujemy wtedy oznaczenie $X_n \xrightarrow{r} X$.

Fakt 11.17:

Jeżeli $r \geq s$, to $X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{s} X$.

Lemat 11.18 (Kroneckera):

Jeżeli (b_n) jest ciągiem monotonicznie rosnącym do nieskończoności oraz $\sum_{i=1}^{\infty} \frac{a_i}{b_i} < \infty$, to

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n a_i}{b_n} = 0.$$

Twierdzenie 11.19 (O zbieżności martyngałów).

Niech (S_n) będzie martyngałem oraz $\mathbb{E}(S_n^2) < \infty$. Istnieje wówczas zmienna losowa S taka, że $S_n \xrightarrow{a.s.} S$ oraz $S_n \xrightarrow{2} S$.

Wniosek 11.20 (Mocne prawo wielkich liczb). Niech X_1, X_2, \dots, X_n będą zmiennymi losowymi niezależnymi o takich samych rozkładach (*iid*). Niech $S_n = X_1 + X_2 + \dots + X_n$. Wtedy

$$\exists \mu \in \mathbb{R} : \frac{S_n}{n} \xrightarrow{a.s.} \mu \Leftrightarrow \mathbb{E}(|X_1|) < \infty.$$

Wówczas $\mu = \mathbb{E}(X_1)$.

Dowód. Załóżmy bez straty ogólności, że $\mu = 0$. Zdefiniujmy zmienną losową

$$S'_n = \sum_{i=1}^n \frac{X_i}{i}.$$

Wtedy

$$S'_{n+1} = S'_n + \frac{X_{n+1}}{n+1}.$$

(S'_n) jest martyngałem względem (X_n) . Z Twierdzenia 11.19 wiemy, że istnieje zmienna losowa S taka, że $S'_n \xrightarrow{a.s.} S$. Stąd dostajemy, że

$$\mathbb{P}\left(\sum_{i=1}^{\infty} \frac{X_i}{i} < \infty\right) = 1.$$

Z Lematu 11.18, podstawiając za $b_n := n$, $a_i := X_i$ otrzymujemy:

$$\sum_{i=1}^n \frac{X_i}{n} = \frac{S_n}{n} \xrightarrow{a.s.} 0$$

Udowodniliśmy zatem implikację w jedną stronę. Pokazanie drugiej implikacji zostało pominięte. \square

Twierdzenie 11.21 (Nierówność Dooba-Kołmogorowa).

Jeżeli (S_n) jest martyngałem względem (X_n) , to dla każdego $\varepsilon > 0$ zachodzi nierówność:

$$\mathbb{P}\left(\max_{1 \leq i \leq n} |S_i| \geq \varepsilon\right) \leq \frac{\mathbb{E}(S_n^2)}{\varepsilon^2}.$$

12. Wykład 12: Martynały: nierówność Hoeffdinga

Wstęp

Na ostatnim wykładzie przedstawiliśmy twierdzenie o zbieżności martynałów oraz nierówność Dooba-Kołmogorowa. Nierówność ta jest wykorzystywana w dowodzie twierdzenia o zbieżności. Na początku zatem przedstawimy jej dowód, a następnie przejdziemy do dowodu twierdzenia o zbieżności martynałów. Na końcu przedstawimy klasyczną nierówność Hoeffdinga. Tym sposobem zakończymy wykład z Rachunku Prawdopodobieństwa 2 w semestrze zimowym 2008/2009 na Wydziale Matematyki i Informatyki UAM w Poznaniu.

12.1. Dowody twierdzeń z poprzedniego wykładu

Dowód Twierdzenia 11.21. Weźmy ciąg zdarzeń (A_n) taki, że $A_0 = \Omega$ oraz dla $k \geq 1$:

$$A_k = \left\{ \max_{i=1, \dots, k} \{|S_i|\} < \varepsilon \right\}$$

Czyli A_k oznacza zdarzenie, że w ciągu (S_n) do k -tego kroku nie przekroczymy ε . Zdefiniujmy zdarzenia (dla każdego $k \geq 1$)

$$B_k = A_{k-1} \cup \{|S_k| \geq \varepsilon\}$$

Czyli B_k oznacza, że po raz pierwszy ε osiągniemy lub przekroczymy w momencie k . Wówczas oczywiście:

$$\forall k \geq 0 : A_k \cup \bigcup_{i=1}^k B_i = \Omega$$

co w języku indykatorów jest równoważne:

$$I_{A_k} + \sum_{i=1}^k I_{B_i} = 1$$

Oszacujmy z dołu wartość oczekiwaną stojącą po prawej stronie dowodzonej nierówności

$$\mathbb{E}(S_n^2) \geq \sum_{i=1}^n \mathbb{E}(S_n^2 I_{B_i}) + \mathbb{E}(S_n^2 I_{A_n}) \geq \sum_{i=1}^n \mathbb{E}(S_n^2 I_{B_i})$$

$$\begin{aligned} \mathbb{E}(S_n^2 I_{B_i}) &= \mathbb{E}((S_n - S_i + S_i)^2 I_{B_i}) = \\ &= \mathbb{E}((S_n - S_i)^2 I_{B_i}) + 2\mathbb{E}((S_n - S_i)S_i I_{B_i}) + \mathbb{E}(S_i^2 I_{B_i}) \end{aligned}$$

Z zadania domowego 10.6 wiadomo, że druga część sumy jest równa zero, a pierwsza jest nieujemna. Natomiast ostatni wyraz możemy oszacować z dołu

$$\mathbb{E}(S_i^2 I_{B_i}) = \mathbb{E}(S_i^2 I_{B_i} | B_i) \mathbb{P}(B_i) + \mathbb{E}(S_i^2 I_{B_i} | B_i^c) \mathbb{P}(B_i^c) = \mathbb{E}(S_i^2 I_{B_i} | B_i) \mathbb{P}(B_i) \geq \varepsilon^2 \mathbb{P}(B_i)$$

Zatem ostatecznie otrzymujemy

$$\mathbb{E}(S_n^2) \geq \sum_{i=1}^n \varepsilon^2 \mathbb{P}(B_i) = \varepsilon^2 \mathbb{P}\left(\max_{i=1, \dots, n} \{|S_i|\} \geq \varepsilon\right)$$

Co należało dowieść. □

Teraz możemy przejść już do dowodu właściwego twierdzenia o zbieżności martyngałów.

Dowód Twierdzenia 11.19. Z zadania domowego 10.7 wynika istnienie takiego ciągu. Wystarczy więc pokazać zbieżność. Pokażemy, że ciąg $(S_n(\omega))$ jest prawie na pewno (a.s.) ciągiem Cauchy'ego, tzn. dla pewnego zdarzenia C zachodzi $\mathbb{P}(C) = 1$. Zdefiniujmy C następująco:

$$C = \{\omega : S_m(\omega) - S_n(\omega) \rightarrow 0; m, n \rightarrow 0\}$$

Zbiór C możemy zapisać równoważnie

$$\begin{aligned} C &= \{\omega : \forall \varepsilon > 0 \exists m : |S_{m+i} - S_{m+j}| < \varepsilon; i, j \geq 1\} = \\ &= \{\omega : \forall \varepsilon > 0 \exists m : |S_{m+i} - S_m| < \varepsilon, i \geq 1\} = \\ &= \left\{ \omega : \forall k |S_{m+i} - S_m| < \frac{1}{k} \right\} \end{aligned}$$

Oznaczmy przez $Y_i = |S_{m+i} - S_m|$. Określmy dopełnienie zbioru C w następujący sposób:

$$C^c = \bigcup_k \bigcap_m A_m(k)$$

gdzie

$$A_m(k) = \bigcup_{n=1}^{\infty} \left\{ \exists i : Y_i \geq \frac{1}{k} \right\}$$

a więc $\bigcap_m A_m(k)$ jest wstępującym ciągiem zdarzeń, które zawierają te ciągi (S_n) , które od m tego miejsca po pewnych i krokach zmieniają się o więcej niż $1/k$. Widać, że

$$\mathbb{P}(C^c) = \lim_{k \rightarrow \infty} \mathbb{P}\left(\bigcap_m A_m(k)\right) \leq \lim_{k \rightarrow \infty} \lim_{m \rightarrow \infty} \mathbb{P}(A_m(k))$$

Oznaczmy przez

$$D_m(k) = \lim_{m \rightarrow \infty} \mathbb{P}(A_m(k)).$$

Naszym celem jest pokazać, że dla każdego k zachodzi $D_m(k) = 0$. Z założenia (S_n) jest martyngałem. Ustalmy teraz m i rozpatrzmy ciąg (Y_n) ($Y_n = S_{m+n} - S_m$). (Y_n) jest martyngałem względem samego siebie, ponieważ

$$\begin{aligned} \mathbb{E}(Y_{n+1} | Y_1, \dots, Y_n) &= \mathbb{E}(\mathbb{E}(Y_{n+1} | S_1, \dots, S_{m+n}) | Y_1, \dots, Y_n) = \\ &= \mathbb{E}(\mathbb{E}(S_{m+n+1} - S_m | S_1, \dots, S_{m+n}) | Y_1, \dots, Y_n) = \\ &= \mathbb{E}(S_{m+n} - S_m | Y_1, \dots, Y_n) = \\ &= \mathbb{E}(Y_n | Y_1, \dots, Y_n) = Y_n \end{aligned}$$

Teraz możemy zastosować Twierdzenie 11.21.

$$\mathbb{P}\left(\max_{i=1,\dots,k}\{|Y_i|\} \geq \frac{1}{k}\right) \leq k^2 \mathbb{E}(Y_n^2)$$

Dokonajmy następującego oszacowania z góry

$$\mathbb{P}(A_m(k)) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\max_{i=1,\dots,k}\{|Y_i|\} \geq \frac{1}{k}\right) \leq k^2 \lim_{n \rightarrow \infty} \mathbb{E}(S_{n+m} - S_m)^2 \quad (12.1.1)$$

Dla każdego m i n zmienne losowe S_m i $S_{m+n} - S_m$ są nieskorelowane, czyli wariancja „udaje”, że jest liniowa, co oznacza:

$$\mathbb{E}(S_m(S_{n+m} - S_m)) = 0$$

oraz

$$\mathbb{E}(S_{n+m})^2 = \mathbb{E}(S_m)^2 + \mathbb{E}(S_{n+m} - S_m)^2 \quad (12.1.2)$$

Wzór 12.1.2 mówi o tym, że ciąg $(\mathbb{E}(S_m^2))$ jest niemalejący i ograniczony, a zatem jest on zbieżny. Jest zatem również ciągiem Cauchy’ego i jego podwójna granica dąży do zera. Wstawiając wzór 12.1.2 do wzoru 12.1.1 otrzymujemy

$$\mathbb{P}(A_m(k)) \leq k^2 \lim_{n \rightarrow \infty} [\mathbb{E}(S_{n+m}^2) - \mathbb{E}(S_m^2)]$$

a więc

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_m(k)) \leq k^2 \cdot 0 = 0$$

Co należało dowieść. □

Przykład 12.1. Kontynuując przykład 11.14 część (a) policzmy:

$$\mathbb{E}(W_n^2) = 1 + \frac{\sigma^2(1 - \mu^{-n})}{\mu(\mu - 1)} = O(1)$$

Zatem (W_n) jest martyngałem względem (Z_n) , a jego drugi moment zwykły jest skończony. Na podstawie Twierdzenia 11.19 istnieje zmienna losowa W taka, że $W_n \xrightarrow{a.s.} W$, a ponadto dla każdego $w > 0$

$$\mathbb{P}(Z_n \geq w\mu^n) \xrightarrow{a.s.} \mathbb{P}(W \geq w)$$

12.2. Nierówność Hoeffdinga

Definicja 12.2. Niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie przestrzenią probabilistyczną. Wówczas ciąg σ -algebr (\mathcal{G}_n)

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \dots \subseteq \mathcal{G}_n \subseteq \dots \subseteq \mathcal{F}$$

nazywamy *filtracją*.

Przykład 12.3. Niech Y będzie zmienną losową $Y : \Omega \rightarrow \mathbb{R}$ o skończonym drugim momencie zwykłym, a (\mathcal{G}_n) filtracją. Dla tak określonej zmiennej losowej zdefiniujemy *martyngał Dooba*:

$$Y_n = \mathbb{E}(Y | \mathcal{G}_n)$$

Pokażemy że (Y_n) jest martyngałem względem samego siebie. Dla każdego $G \in \mathcal{G}_n$ zachodzi

$$\begin{aligned} 0 &= \mathbb{E}((Y - Y_n)I_G) = \mathbb{E}((Y - Y_{n+1} + Y_{n+1} - Y_n)I_G) = \\ &= \mathbb{E}((Y - Y_{n+1})I_G) + \mathbb{E}((Y_{n+1} - Y_n)I_G) \stackrel{\text{definicja}}{=} \\ &= \mathbb{E}((Y_{n+1} - Y_n)I_G) \end{aligned}$$

Zatem

$$Y_n = \mathbb{E}(Y_{n+1} | \mathcal{G}_n)$$

Wniosek 12.4. Niech (\mathcal{F}_n) będzie filtracją, natomiast Y_n będzie \mathcal{F}_n -mierzalne. Wówczas para (Y_n, \mathcal{F}_n) będzie martyngałem, jeśli dla każdego n zachodzą następujące warunki:

- (a) $\mathbb{E}(Y_n) < \infty$
- (b) $\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = Y_n$

Przedstawmy jeszcze krótki fakt:

Fakt 12.5:

Jeżeli (Y_n) jest martyngałem względem (X_n) , to (Y_n) jest martyngałem względem $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$

Definicja 12.6. *Różnicą martyngałów* nazywamy funkcję D_n zmiennych losowych taką, że $D_n = Y_n - Y_{n-1}$, czyli

$$Y_n = Y_0 + \sum_{i=1}^n D_i$$

Fakt 12.7:

D_n jest \mathcal{F}_n -mierzalne i ponadto

- (a) $\mathbb{E}(|D_n|) < \infty$
- (b) $\mathbb{E}(D_{n+1} | \mathcal{F}_n) = 0$

Teraz możemy przejść do właściwego twierdzenia, któremu jest poświęcony ten rozdział.

Twierdzenie 12.8 (Twierdzenie Hoeffdinga).

Niech (Y_n, \mathcal{F}_n) będzie martyngałem. Załóżmy, że istnieje ciąg liczbowy K_1, K_2, \dots niekoniecznie o skończonych wartościach, taki że dla dowolnego n zachodzi $\mathbb{P}(|D_n| \leq K_n) = 1$. Wówczas

$$\mathbb{P}(|Y_n - Y_0| \geq x) \leq 2 \exp \left\{ -\frac{x^2}{2 \sum_{i=1}^n K_i^2} \right\}$$

Dowód. W dowodzie należy wykorzystać nierówność Markowa biorąc wykładnicze zmienne losowe i dokonując odpowiedniej optymalizacji. Z powodu braku czasu dowód zostawiamy jako zadanie domowe dla chętnych. \square

Przykład 12.9. Niech (X_n) będzie ciągiem niezależnych zmiennych losowych o rozkładzie Bernoulliego z parametrem p (*iid*). Wówczas zmienna losowa S_n będąca sumą tych zmiennych losowych ma rozkład dwumianowy z parametrami n i p

$$S_n = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$$

Wówczas ciąg (Y_n) ($Y_n = S_n - \mathbb{E}(S_n) = S_n - np$) jest martyngałem, gdzie $S_0 = Y_0 = 0$. Ponadto $D_n = Y_n - Y_{n-1} = X_n - p \in \{-p, 1-p\}$ a więc $|D_n| \leq 1 = K_n$

Na podstawie Twierdzenia 12.8 mamy

$$\mathbb{P}(|S_n - np| \geq x) \leq 2 \exp \left\{ -\frac{x^2}{2n} \right\}$$

Przyjmując $x = \log(n)$ otrzymamy oszacowanie prawdopodobieństwa lewej strony z góry przez liczbę 2. Jednak takie oszacowanie jest całkowicie bezużyteczne. Natomiast podstawienie $x = a_n \sqrt{n}$ daje dużo lepsze oszacowanie, o ile ciąg (a_n) ma niezerową granicę. Dzięki temu możemy otrzymać nierówność Chernoffa:

$$\mathbb{P}(|S_n - np| \geq a_n \sqrt{n}) \leq 2 \exp \left\{ -\frac{na_n^2}{2n} \right\} = 2 \exp \left\{ -\frac{a_n^2}{2} \right\}$$

13. Appendix

A - Zadania domowe

Tutaj można by było umieścić zadania domowe.